PRODUCTION OF 2'-DEOXYNUCLEOSIDES AND 2'-DEOXYNUCLEOSIDE PRECURSORS FROM 2-DEHYDRO-3-DEOXY-D-GLUCONATE

This invention relates to a process for preparing 2'-deoxynucleoside compounds or 2'-deoxynucleoside precursors using 2-dehydro-3-deoxy-D-gluconic acid (usually abbreviated as KDG) or its salts as a starting material. A variety of 2'-deoxynucleosides and their analogues are used as a starting material for synthesis or drug formulation in production of an antiviral, anticancer or antisense agent.

Specifically, the invention relates to a method in which KDG or a derivative of KDG is subjected to a decarboxylation step to remove the original carboxy group of KDG. In a preferred embodiment, the KDG used in the method according to the invention is enzymatically produced from D-gluconate or D-glucosaminate.

2'-deoxynucleosides and 2'-deoxynucleoside precursors including 2-deoxy-D-ribose are used as starting material for synthesis or drug formulation, for instance, in production of antiviral and anticancer agent. 2'-deoxynucleosides or derivatives thereof and 2'-deoxynucleoside precursors are also used as reagents for research, diagnosis and synthesis of therapeutic antisense molecules.

In one method of the prior art, deoxynucleosides are generated from biological materials such as testis (WO 99/49074) or yeast or fish sperm by enzymatic cleavage of DNA. This method, however, involves several disadvantages, in particular regarding difficulties of obtaining the starting material in sufficient quantity and quality.

The main production process of 2-deoxy-D-ribose currently consists in chemical hydrolysis of DNA. In this case, the deoxyribosyl moiety originates in ribonucleotide reductase activity. No synthesis of 2-deoxy-D-ribose from KDG has been yet described.

In most living cells, deoxyribonucleosides result from a "salvage pathway" of the

nucleotide metabolism. The deoxyribose moiety of deoxyribonucleosides is obtained through the reduction of a ribosyl moiety into di- or triphosphate ribonucleotides catalyzed by ribonucleotide reductases. However, the deoxyribose moiety is not recycled, but is degraded into D-glyceraldehyde-3-phosphate and acetaldehyde following the reactions of central metabolism:

- deoxynucleoside is cleaved into deoxyribose-1-phosphate and nucleobase through phosphorolysis mediated by products of the genes encoding thymidine phosphorylase (deoA), purine-nucleoside phosphorylase (deoD), uridine phosphorylase (udp) or xanthosine phosphorylase (xapA).
- deoxyribose-1-phosphate is converted into deoxyribose-5-phosphate through a reaction catalyzed by deoxyribose phosphate mutase (deoB),
- which is further degraded to D-glyceraldehyde-3-phosphate and acetaldehyde through a reaction catalyzed by deoxyribose-5-phosphate aldolase (deoC).

It has been shown that the deo enzymes also catalyze in vitro the reverse anabolic reactions: Deoxyribose-5-phosphate is obtained in vitro in the presence of purified Escherichia coli or Lactobacillus plantarum deoxyribose aldolase starting from acetaldehyde and D-glyceraldehyde-3-phosphate (Rosen et al., J. Biol. Chem., 240, (1964), 1517-1524; Pricer, J. Biol. Chem., 235, (1960), 1292-1298). Deoxyribose can also be obtained with acetaldehyde and glyceraldehyde as enzyme substrates, but only with a very low yield (Barbas, J. Am. Chem. Soc. 112 (1990), 2013-2014).

The patent application WO 01/14566 describes the enzymatic synthesis of deoxynucleosides starting from deoxyribose-1-phosphate through the combined activities of three enzymes of the deo operon, i.e. deoxyribose aldolase, deoxyribomutase and phosphorylase (thymidine or purine nucleoside phosphorylase) in a one-pot reaction, using as starting substrates glyceraldehyde-3-phosphate, acetaldehyde and a nucleobase. D-glyceraldehyde-3-phosphate can be obtained from fructose-1,6-bisphosphate by an enzymatic process.

The patent application EP 1179598 describes the use of phosphorylase to catalyze the enzymatic production of deoxynucleosides starting from deoxyribose-1-phosphate and nucleobase. The yield of deoxynucleoside synthesis is improved by

precipitation of phosphate.

However, methods using enzymes of the deo operon working in the reverse direction compared to their biological function show low yields, which indicates serious drawbacks for their use.

In view of the above-described ineffectiveness of the currently applied processes for producing deoxynucleosides and deoxynucleoside precursors, it is an object of the present invention to provide means and methods for the biosynthetic production of deoxynucleosides and deoxynucleoside precursors starting from cheap and commercially available compounds without being dependent on unreliable natural sources.

In particular, there is a need for alternative methods for the production of deoxynucleosides and deoxynucleoside precursors which allow efficient and economical synthesis of deoxyribonucleosides, by means of which the drawbacks of prior art processes are eliminated.

The present invention relates to a method for producing 2'-deoxynucleosides and precursors thereof starting from 2-dehydro-3-deoxy-D-gluconic acid (KDG) or its salts and comprising a decarboxylation step.

In particular, this method is useful for producing 2-deoxy-D-ribose (DRI) as well as synthetically versatile enamine derivatives of DRI as 2'-deoxynucleoside precursors.

The decarboxylation step takes place by reacting either KDG or its salts directly, or a derivative of KDG, usually to cleave the C1-C2 bond of the KDG.

In one embodiment of the invention, KDG or one of its salts undergoes (oxidative) decarboxylation leading to 2-deoxy-D-ribonic acid (DRN) or its salts, itself being further converted into 2-deoxy-D-ribose (DRI) or 2-deoxy-D-ribitol (DRL).

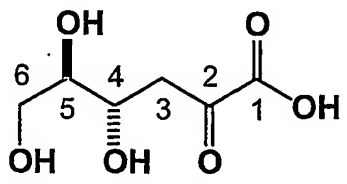
In another embodiment of the invention, decarboxylation takes place by reacting KDG or its salts with an amine, leading to an enamine derivative. This high energy enamine derivative can be further converted into DRI by hydrolysis.

In another embodiment of the invention, (oxidative) decarboxylation is carried out on 3-deoxy-D-gluconic acid (DGN) or its salts and/or 3-deoxy-D-mannonic acid (DMN) or its salts as derivatives of KDG, leading to DRI. Production of a mixture of DGN and DMN takes place by reduction of KDG. The decarboxylation is preferably carried out via reaction with hydrogen peroxide.

In another embodiment of the invention, (oxidative) decarboxylation is carried out on 3-deoxy-D-glucosaminic acid (DGM) or its salts and/or 3-deoxy-D-mannosaminic acid (DMM) or its salts, leading to DRI. Production of a mixture of DGM and DMM takes place from KDG by reductive amination.

Another aspect of the invention is a convenient and cost-effective method for preparing KDG or its salts to be used in the above methods. This method starts either from D-gluconate or from D-glucosaminate through the use of recombinant enzymes. The invention provides a novel nucleotide sequence encoding a polypeptide having D-gluconate dehydratase activity and a nucleotide sequence encoding a polypeptide having D-glucosaminate deaminase activity.

The starting material used for the method of the present invention is KDG, represented by formula (I) below or one of its salts, or a protected derivative thereof wherein one ore more of the hydroxyl groups at positions 4, 5 and/or 6 are protected by a protection group known in the art.



(I)

The term "2'-deoxynucleoside" as used herein relates to 2'-deoxyribonucleosides which are N-glycosides, and wherein the basic N-atom of the nucleobase or nucleobase analog is bound to the anomeric carbon atom of 2-deoxy-D-ribose, or one of its derivatives. Examples of a suitable nucleobase are adenine, cytosine, guanine, thymine, uracil, 2,6-diaminopurine, and hypoxanthine. Examples of nucleobase analogs are 5-azacytosine, 2-chloro-adenine, 5-iodo-cytosine, 8-aza-guanine, 5-iodo-uracil, 5-bromo-uracil, 5-fluoro-uracil, 5-ethyl-uracil and 5-trifluoromethyl-uracil.

The term "2'-deoxynucleoside precursors" as used herein, relates to compounds which can be easily converted into 2'-deoxynucleosides by applying methods known in the prior art. Preferred 2'-deoxynucleoside precursors are 2-deoxy-D-ribose (DRI) or carbohydrate compounds which can be converted into the 2-deoxy-D-ribosyl moiety of 2'-deoxynucleosides, for instance, those established in the prior art 1-phospho-2-deoxy-D-ribose, 5-phospho-2-deoxy-D-ribose and those established by the present invention 2-deoxy-D-ribitol, 2-deoxy-D-ribonic acid, 2-deoxy-D-ribono-1,4-lactone, 1-N-morpholino-3,4,5-trihydroxy-pentene-1, and their derivatives.

The method of the invention encompasses methods wherein the decarboxylation step is directly carried out on KDG or its salts or on compounds derived from KDG. Preferred KDG derivatives are 3-deoxy-D-gluconic acid, 3-deoxy-D-mannonic acid, 3-deoxy-D-glucosaminic acid and 3-deoxy-D-mannosaminic acid and their respective salts.

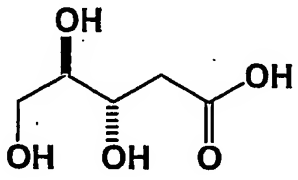
Furthermore, KDG and its salts or protected forms of these wherein one or more of the hydroxyl groups at the positions 4,5 and /or 6 are replaced by protecting groups known for that purpose in the art are also suitable starting materials for the decarboxylation reaction of the present invention. Unless noted otherwise, any reference to KDG in the following specification embraces protected forms of KDG, just as reference to KDG derivatives is intended to embrace protected forms of these derivatives. Similarly, any reference to the products obtained in the methods of the invention is intended to encompass protected forms of these products. Preferred

protection groups for the purpose of the invention are those which replace the respective hydroxyl groups by acetate ester, benzoate ester, allyl ether, benzyl ether, trityl ether, ter-butyldimethylsilyl (TBDMS) ether, isopropylidene or a benzylidene acetal.

It should be understood that, depending on suitable reaction conditions for the embodiments of the invention, the carboxylic groups contained in the organic acids used as reactants or obtained as products can be in a protonated form or in their salt form, or may be present in equilibrium. Exemplary salts of these acids are those which have metal or ammonium ions as counterions, particularly alkali metal ions such as sodium and/or potassium.

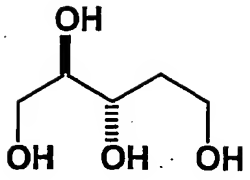
Most of the carbohydrate compounds and their derivatives described in the present invention exist under several cyclic form but for simplicity reasons have been represented by open chain formulas. It is understood that the present invention encompasses all these isomeric or tautomeric forms.

In a first embodiment of the invention, KDG or its salts is reacted with hydrogen peroxide and undergoes (oxidative) decarboxylation to 2-deoxy-D-ribonic acid (DRN), a compound of formula (II) or its salts.



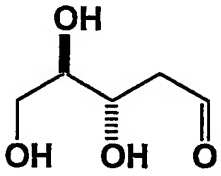
(II)

The product may be further converted into or 2-deoxy-D-ribitol (DRL), represented by formula (IV)



(IV)

or 2-deoxy-D-ribose (DRI), represented by formula (III)



(III)

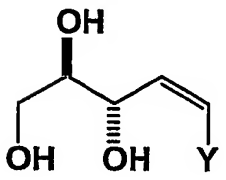
DRN, DRL and particularly DRI are among preferred 2'-deoxynucleoside precursors for the purpose of the present invention. Conversion of DRN to DRI may proceed directly or via DRL as an intermediate.

Preferably, the preparation of DRN is carried out by oxidative decarboxylation of sodium or potassium 2-dehydro-3-deoxy-D-gluconate in aqueous solution with hydrogen peroxide at room temperature as described in example 5. A general method for the preparation of aldonic acids by oxidative decarboxylation of 2-ketoaldonic acids is described in patent EP 1 038 860 A1.

Preferably, the preparation of DRL is carried out by hydrogenation of 2-deoxy-D-ribonolactone in aqueous solution with Rhodium catalyst on carbon at a temperature of 130°C under a pressure of 80 bars as described in example 6. 2-Deoxy-D-ribonolactone can be easily prepared by converting a 2-deoxy-D-ribonate (DRN salt) into 2-deoxy-D-ribonic acid, which is in equilibrium with its lactonic form in aqueous solutions (Han, Tetrahedron. 1993. 49, 349-362; Han, Tetrahedron Asymmetry. 1994. 5, 2535-62).

Preferably the preparation of 2-deoxy-D-ribose (DRI) is carried out by oxidization of 2-deoxy-D-ribitol (DLR), e.g. with chromium oxide in pyridine.

In another embodiment of the invention, decarboxylation takes place by reacting (KDG) or its salts with an amino group-containing reagent Y-H leading to a compound of formula (V).



(V)

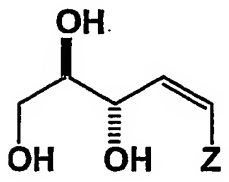
or its respective trans isomer or a protected form thereof, as a 2'-deoxynucleoside precursor. Y-H represents an amine with the hydrogen atom H bound to the nitrogen of the amino group.

In a preferred embodiment of the invention, the amino group-containing reagent represented by Y-H is a linear or cyclic secondary amine; a primary amine that possess a β-carbonyl group, preferably 3-amino-2-indolinone which was found to be effective for the decarboxylation of α-keto acids (Hanson, J. Chem. Education, 1987, 591-595). In each of these cases, -Y in formula (V) represents the respective nitrogen containing residue derived from these amino-group containing reagent.

Preferably, the compound of formula (V) represents an enamine produced via reaction of a linear or cyclic secondary amine as Y-H.

Preferred cyclic secondary amines are morpholine, pyrrolidine, piperidine, or N-methyl piperazine; preferred non-cyclic amines are those of the formula R_1-NH-R_2, wherein R_1 and R_2 independently represent a linear or branched alkyl group of 1-8, preferably 1 to 4 carbon atoms. Particularly preferred as a non-cyclic amine is diethylamine.
Particularly preferred as a cyclic amine is morpholine.

The compound of formula (V) or its trans isomer or a protected form thereof can be further reacted with Z-H, wherein H represents a hydrogen atom and Z represents a leaving group, to produce a compound of formula (VI)



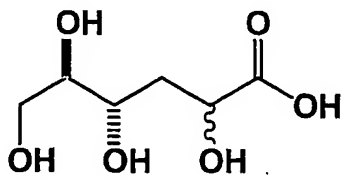
(VI)

or its respective trans isomer or a protected form thereof, as a 2'-deoxynucleoside precursor. Z-H is preferably water, in which case the compound of formula (VI) is DRI or a protected form thereof (keto-enol-tautomerism).

Preferably, the preparation of the compound of formula (V) is carried out by reacting KDG in benzene with the amine, e.g. morpholine under reflux using the method described in example 7, leading to 1-N-morpholino-3,4,5-trihydroxy-pentene-1. Acid catalysed hydrolysis with water yields 2-deoxy-D-ribose (DRI)

A general route to aldehydes via enamines from α–oxocarboxylic acids carrying β–hydrogens is described by Stamos (Tetrahedron Lett. 23 (1982), 459-462). Other methods for the preparation and hydrolysis of enamines have been described elsewhere (Stork, J. Am. Chem. Soc. 85 (1963), 207-222; Stamhuis, J. Org. Chem. 30 (1965), 2156-2160).

In another embodiment of the invention, KDG or its salt is converted to 3-deoxy-D-gluconic acid (DGN) and/or 3-deoxy-D-mannonic acid (DMN) represented by formula (VII) or the salts of these compounds



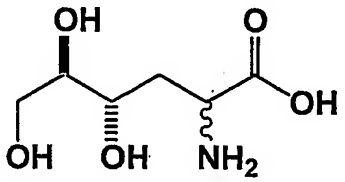
(VII)

The products resulting from this reaction undergo (oxidative) decarboxylation, preferably using hydrogen peroxide, to yield DRI. Production of a mixture of DGN and DMN or their salts takes place from KDG or its salts by reduction.

Preferably the preparation of 2-deoxy-D-ribose (DRI) is carried out by non-stereoselective reduction of 2-dehydro-3-deoxy-D-gluconic acid in water with sodium borohydride at room temperature using the method described for 2-keto-3-deoxyheptonic acid by Weissbach (J. Biol. Chem. 234 (1959), 705-709), followed by oxidative decarboxylation of 3-deoxy-D-gluconate and 3-deoxy-D-mannonate with hydrogen peroxide as described e.g. in US patent 3,312,683; Richards J. Chem. Soc. (1954), 3638-3640; Sowden J. Am. Chem. Soc. 76 (1954), 3541-3542.

In another preferred embodiment, the preparation of a mixture of DGN and DMN is carried out by hydrogenation of 2-dehydro-3-deoxy-D-gluconate in aqueous solution with 6% mol Nickel Raney catalyst or Platinum oxide at room temperature under a pressure of 6 bars.

In another embodiment of the invention, KDG or its salt is converted to 3-deoxy-D-glucosaminate (DGM) or 3-deoxy-D-mannosaminate (DMM) represented by formula (VIII) or the salts of these compounds



(VIII)

The products resulting from this reaction undergo (oxidative) decarboxylation, preferably using ninhydrin, to yield DRI. Production of a mixture of DGM and DMM or their salts takes place from KDG or its salts by reductive amination.

Preferably the preparation of 2-deoxy-D-ribose is carried out by non-stereoselective reductive amination of sodium or potassium 2-dehydro-3-deoxy-D-gluconate in

aqueous solution with ammonia and sodium cyanoborohydride at room temperature, followed by oxidative decarboxylation of 3-deoxy-D-2-glucosaminate and 3-deoxy-D-2-mannosaminate with ninhydrin using the method described for the synthesis of 2-deoxy-D-allose by Shelton (J. Am. Chem. Soc. 118 (1996), 2117-2125; and Borch, J. Am. Chem. Soc. 93 (1971), 2897; Durrwachter, J. Am. Chem. Soc. 108 (1986), 7812 referenced therein).

Furthermore, the present invention provides a method for producing the compound of formula (III) (2-deoxy-D-ribose) by converting the compound of formula (I) or one of its salts (KDG) in one single step. Preferably this conversion is achieved through enzymatic catalysis. This conversion is preferably catalysed by a keto acid decarboxylase. Preferred keto acid decarboxylases are thiamin pyrophosphate (TPP) dependent keto acid decarboxylases. Examples of TPP dependent keto acid decarboxylases are pyruvate decarboxylase (EC 4.1.1.1), a benzoylformate decarboxylase (EC 4.1.1.7), an indolepyruvate decarboxylase (EC 4.1.1.74), a phosphonopyruvate decarboxylase, a sulfopyruvate decarboxylase (EC 4.1.1.79), an oxalyl-coenzymeA decarboxylase (EC 4.1.1.8), an oxoglutarate decarboxylase (EC 4.1.1.71) or a phenylpyruvate decarboxylase (EC 4.1.1.43). It could be shown that keto acid decarboxylases, e.g., pyruvate decarboxylase enzymes from different organisms, can convert KDG into 2-deoxy-D-ribose (see Examples 8 to 12). In principle any keto acid decarboxylase can be used in connection with the present invention.

In a preferred embodiment of the method according to the invention KDG is converted into 2-deoxy-D-ribose by use of an enzyme having pyruvate decarboxylase activity.

A pyruvate decarboxylase catalyses the following reaction:

$$\text{pyruvate} + H^+ \rightarrow \text{acetaldehyde} + CO_2$$

Several pyruvate decarboxylases (PDC) have been characterized as well as the corresponding pdc genes, for instance PDC from Zymomonas mobilis (Genbank accession number AAD19711; Neale et al., J. Bacteriol. 1987, 169:1024-1028), PDC from Saccharomyces cerevisiae (Genbank accession number NP013145; Candy et al., J. Gen. Microbiol. 1991, 137:2811-2815), PDC from Acetobacter pasteurianus (Genbank accession number AAM21208; Raj et al., Arch. Microbiol. 2001, 176:443-451), PDC from Zymobacter palmae (Genbank accession number AAM49566; Raj et al., Appl. Environ. Microbiol. 2002, 68:2869-2876), PDC from Sarcina ventriculi (Genbank accession number AAL18557; Lowe et al., J. Gen. Microbiol. 1992, 138:803-807). Many other pyruvate decarboxylases seems to occur in plants, fungi and bacteria as evidenced by the occurrence in these organisms of genes sharing sequence homologies with well-established pdc genes. Examples of such putative pyruvate decarboxylases are:

PDC from plants:

Arabidopsis thaliana (Genbank accession number T48155)
Echinochloa crus-galli (Genbank accession number AAM18119)
Oryza sativa (Genbank accession number NP922014)
Rhizopus oryzae (Genbank accession number AAM73540)
Lotus corniculatus (Genbank accession number AAO72533)
Zea mays (Genbank accession number BAA03354)
Pisum sativum (Genbank accession number CAA91445)
Garden pea (Genbank accession number S65470)
Nicotiana tabaccum (Genbank accession number CAA57447)
Solanum tuberosum (Genbank accession number BAC23043)
Fragaria ananassa (Genbank accession number AAL37492)
Cucumis melo (Genbank accession number AAL33553)
Vitis vinifera (Genbank accession number AAG22488)

PDC from Fungi :

Saccharum officinarum (Genbank accession number CAB61763)
Aspergillus orizae (Genbank accession number AAD16178)
Aspergillus parasiticus (Genbank accession number P51844)
Saccharomyces cerevisiae (Genbank accession number NP013145)
Flammulina velutipes (Genbank accession number AAR00231)

Saccharomyces kluyveri (Genbank accession number AAP75899)
Schizosaccharomyces pombe (Genbank accession number CAB75873)
Candida glabrata (Genbank accession number AAN77243)
Neurospora crassa (Genbank accession number JN0782)
Pichia stipis (Genbank accession number AAC03164)
Kuyveromyces lactis (Genbank accession number CAA61155)
Emericella nidulans (Genbank accession number AAB63012)

PDC from Prokaryotes:

Mycobacterium bovis (Genbank accession number CAD93738)
Mycobacterium leprae (Genbank accession number CAC31122)
Mycobacterium tuberculosis (Genbank accession number NP215368)
Mycoplasma penetrans (Genbank accession number NP758077)
Clostridium acetobutylicum (Genbank accession number NP149189)
Acetobacter pasteurianus (Genbank accession number AAM21208)
Zymobacter palmae (Genbank accession number AAM49566)
Zymomonas mobilis (Genbank accession number AAD19711)
Sarcina ventriculi (Genbank accession number AAL18557)
Nostoc punctiforme (Genbank accession number ZP00110850)

Such enzymes can be easily produced by recombinant microorganisms overexpressing the corresponding gene. Examples of genes coding for TPP dependent keto acid decarboxylases are pdc from Zymomonas mobilis (Genbank accession number AF124349), pdc from Saccharomyces cerevisiae (Genbank accession number NC001144), pdc from Acetobacter pasteurianus (Genbank accession number AF368435), pdc from Zymobacter palmae (Genbank accession number AF474145), pdc from Sarcina ventriculi (Genbank accession number AF354297). Other pdc genes can be found at Genbank corresponding to the above list of putative pyruvate decarboxylases.

In a preferred embodiment the pyruvate decarboxylase is of eukaryotic origin, more preferably it is from yeast and most preferably it is from Saccharomyces cerevisiae. In a particularly preferred embodiment the pyruvate decarboxylase is the pyruvate decarboxylase from S. cerevisiae which has the amino acid sequence as shown in SEQ ID NO: 21 (see also GenBank accession number NP013145).

In another preferred embodiment the pyruvate decarboxylase is of prokaryotic origin, more preferably it is from an organism of the genus Zymomonas and most preferably from Zymomonas mobilis. In a particularly preferred embodiment the pyruvate decarboxylase is the pyruvate decarboxylase from Z. mobilis which has the amino acid sequence as shown in SEQ ID NO: 19 (see also GenBank accession number AAD19711).

In another preferred embodiment the prokaryotic pyruvate decarboxylase is from an organism of the genus Acetobacter, more preferably from the species Acetobacter pasteurianus. Particularly preferred the pyruvate decarboxylase is that of A. pasteurianus which shows the amino acid sequence as given in SEQ ID NO: 25 (see also GenBank accession number AAM21208).

In a further preferred embodiment the pyruvate decarboxylase is from an organism of the genus Zymobacter, more preferably of the species Zymobacter palmae. Particularly preferred is a pyruvate decarboxylase from Z. palmae which shows the amino acid sequence given in SEQ ID NO: 29 (see also GenBank accession number AAM49566).

In another preferred embodiment of the method according to the invention KDG is converted into 2-deoxy-D-ribose by use of an enzyme having benzoylformate decarboxylase activity.

A benzoylformate decarboxylase catalyses the following reaction:

$$\text{benzoylformate} + H^+ \rightarrow \text{benzaldehyde} + CO_2$$

A benzoylformate decarboxylase (BDC) from Pseudomonas putida (Genbank accessing number AAC15502; Tsou et al., Biochemistry. 1990, 29:9856-9862) has been characterized as well as the corresponding gene mdlC (Genbank accessing number AY143338). This enzyme has been shown to decarboxylate both D and L isomers of 2-keto-4,5-dihydroxyvalerate into the respective isomers of 3,4-

dihydroxybutanal (Niu et al., J. Am. Chem. Soc. 125 (2003), 12998-12999). Many other benzoylformate decarboxylases seems to occur in bacteria and archaebacteria as evidenced by the occurrence in these organisms of genes sharing sequence homologies with genes coding for well-established BDC. Examples of such putative benzoylformate decarboxylases are:

BDC from bacteria:

Pseudomonas aeruginosa (Genbank accession number NP_253588)
Rhodopseudomonas palustris (Genbank accession number NP_946955)
Streptomyces coelicolor (Genbank accession number NP_631486)
Chromobacterium violaceum (Genbank accession number NP_902771)
Bradyrhizobium japonicum (Genbank accession number NP_774243)

BDC from archaebacteria:

Sulfolobus solfataricus (Genbank accession number NP_343070)
Thermoplasma acidophilum (Genbank accession number NP_393976)
Thermoplasma volcanium (Genbank accession number NP_111716)

Such enzymes can be easily produced by recombinant microorganisms overexpressing the corresponding bdc gene. Such genes can be found at Genbank corresponding to the above list of putative benzoylformate decarboxylases.

Another example for a thiamine dependent decarboxylase which can be used in the method according to the invention is phosphonopyruvate decarboxylase. Several phosphonopyruvate decarboxylases (PPD) have been characterized as well as the corresponding genes, for instance PPD from Bacteroides fragilis (Genbank accession number AAG26466; Zhang et al., J. Biol. Chem. 2003, 278:41302-41308), PPD from Streptomyces wedmorensis (Genbank accession number BAA32496; Nakashita et al., J. Antibiot. 1997, 50:212-219). Many other phosphonopyruvate decarboxylases seem to occur in bacteria as evidenced by the occurrence in these organisms of genes sharing sequence homologies with genes coding for well-established PPD. Examples of such putative phosphonopyruvate decarboxylases are: PPD from Bacteroides thetaiotaomicron (Genbank accession number NP_810632), PPD from Amycolatopsis orientalis (Genbank accession number CAB45023), PPD from Clostridium tetani E88 (Genbank accession number

NP_782297), PPD from Streptomyces viridochromogenes (Genbank accession number CAA74722), PPD from Streptomyces hygroscopicus (Genbank accession number BAA07055), PPD from Streptomyces coelicolor A3 (Genbank accession number NP_733715), Streptomyces rishiriensis (Genbank accession number AAG29796), Bordetella pertussis (Genbank accession number CAE 41214. Such enzymes can be easily produced by recombinant microorganisms overexpressing the corresponding gene.

A further example of a thiamine dependent decarboxylases which can be used in the method according to the present invention is sulfopyruvate decarboxylase. A sulfopyruvate decarboxylases (SPD) from Methanococcus jannaschii (Graupner et al., J. Bacteriol. 2000. 182:4862-4867) consisting of two subunits ComD (Genbank accession number P58415) and ComE (Genbank accession number P58416) has been characterized as well as the corresponding genes. Many other sulfopyruvate decarboxylases seems to occur in archaebacteria and in bacteria as evidenced by the occurrence in these organisms of genes sharing sequence homologies with genes coding for well-established SPD.

Another further example of thiamine dependent decarboxylase which can be used in the method according to the present invention is indolepyruvate decarboxylase. Several indolepyruvate decarboxylases (IPD) have been characterized as well as the corresponding genes, for instance IPD from Enterobacter cloacae (Genbank accession number BAA14242; Scutz et al., 2003, Eur. J. Biochem. 270:2322-2331), IPD from Azospirillum brasilense (Genbank accession number AAC36886; Costacurta et al., Mol. Gen. Genet. 1994, 243:463-472), IPD from Erwinia herbicola (Genbank accession number AAB06571; Brandl et al., Appl. Environ. Microbiol. 1996, 62:4121-4128). Many other indolepyruvate decarboxylases seem to occur in bacteria as evidenced by the occurrence in these organisms of genes sharing sequence homologies with genes coding for well-established IPD.

Still another further example of a thiamine dependent decarboxylases which can be used in the method according to the present invention is phenylpyruvate decarboxylase. A phenylpyruvate decarboxylase from yeast (Genbank accession

number NP010668; Vuralhan et al., Appl. Environ. Microbiol. 2003, 69:4534-41) has been characterized as well as the corresponding gene ARO10 (Genbank accession number NC001136).

In a preferred embodiment of the method according to the invention in which the decarboxylation step is effected by an enymatic reaction, the pH value is regulated by addition of an acid to be between pH 5 and pH 9, preferably between pH 6 and pH 8. In principle, any suitable acid can be used for this purpose. Preferred acids are HCl, H_2SO_4, D-gluconic acid or 2-dehydro-3-deoxy-D-gluconic acid.

Another aspect of the invention is a convenient and cost-effective method for preparing KDG either from D-gluconate (GCN) or from D-glucosaminate through the use of recombinant enzymes.

In a preferred embodiment of the method of the invention, the compound of formula (I) is produced in a preliminary step from a D-gluconate salt by the use of a D-gluconate dehydratase activity. Preferred salts are potassium or sodium D-gluconate. Preferably the D-gluconate dehydratase is encoded by a polynucleotide comprising the nucleotide sequence selected from the group consisting of:

(a) nucleotide sequences encoding a polypeptide comprising the amino acid sequence of SEQ ID N°2;
(b) nucleotide sequences comprising the coding sequence of SEQ ID N°1;
(c) nucleotide sequences encoding a fragment encoded by a nucleotide sequence of (a) or (b);
(d) nucleotide sequences hybridising with a nucleotide sequence of any one of (a) to (c); and
(e) nucleotide sequences which deviate from the nucleoside sequence of (d) as a result of degeneracy of the genetic code.

The enzymatic synthesis of KDG or its salts using D-gluconate dehydratase proceeds according to the following reaction: D-gluconate is converted into KDG by the elimination of one water molecule. The activity of a D-gluconate dehydratase has been characterized in different bacterial species e.g. in Alcaligenes (Kersters,

Methods in Enzymology 42 (1975), 301-304); Clostridium pasteurianum, (Gottschalk, Methods in Enzymology 90 (1982), 283-287); Thermoplasma acidophilum (Budgen, FEBS Letters 196 (1986), 207-210) and Sulfolobus solfataricus (Nicolaus, Biotechnology Letters 8(7) (1986), 497-500). The preferred D-gluconate dehydratase was identified by screening several collection strains for D-gluconate dehydratase activity. The gene encoding a D-gluconate dehydratase, which was designated gcnD was selected from a genomic library of Agrobacterium tumefaciens strain C58, and further inserted in a multi copy vector optimised for expression. It was shown that a crude extract from E. coli cells over-expressing the gcnD gene catalysed the total conversion of D-gluconate into KDG (see Example 2).

In a further preferred embodiment of the method of the invention, the compound of formula (I) is produced in a preliminary step from D-glucosaminate by the use of a D-glucosaminate deaminase activity. Preferably the D-glucosaminate deaminase is encoded by a polynucleotide comprising the nucleotide sequence selected from the group consisting of:

(f) nucleotide sequences encoding a polypeptide comprising the amino acid sequence of SEQ ID N°4;

(g) nucleotide sequences comprising the coding sequence of SEQ ID N°3;

(h) nucleotide sequences encoding a fragment encoded by a nucleotide sequence of (a) or (b);

(i) nucleotide sequences hybridising with a nucleotide sequence of any one of (a) to (c); and

(j) nucleotide sequences which deviate from the nucleoside sequence of (d) as a result of degeneracy of the genetic code.

The enzymatic synthesis of KDG or its salts using D-glucosaminate deaminase proceeds according to the following reaction: D-glucosaminate is converted into KDG by the elimination of one molecule water and one molecule of ammonia. The activity of a D-glucosaminate deaminase has been characterized in different bacterial species e.g. in Pseudomonas fluorescens (Iwamoto, Agric. Biol. Chem. 53 (1989), 2563-2569) Agrobacterium radiobacter (Iwamoto, FEBS Letters 104 (1979), 131-134;

Iwamoto, J. Biochem. 91 (1982), 283-289), and its requirement for Mn^{2+} ion was shown (Iwamoto, Biosci. Biotech. Biochem. 59 (1995), 408-411).

The preferred D-glucosaminate deaminase was identified by screening several collection strains for D-glucosaminate deaminase activity. The gene encoding a D-glucosaminate deaminase, which was designated gmaA was isolated from Agrobacterium tumefaciens strain C58 by cloning a gene annotated as a putative D-serine deaminase. The gmaA gene was further inserted in a multi copy vector optimised for expression. It was shown that a crude extract from E. coli cells over-expressing the gmaA gene catalysed the conversion of D-glucosaminate into KDG (see Example 4).

In a preferred embodiment the present invention relates to a method for producing a compound of formula III, in particular 2-deoxy-D-ribose, starting from D-gluconate or D-glucosaminate by enzymatic reactions which, in a first step, convert D-gluconate or D-glucosaminate into KDG as described above and, in a second step, convert KDG into 2-deoxy-D-ribose as described above.

Thus, the enzymatic conversion of D-gluconate into KDG can be achieved by the use of a D-gluconate dehydratase. The enzymatic conversion of D-glucosaminate into KDG can be achieved by the use of a D-glucosaminate deaminase. With respect to the preferred embodiments the same applies as has already been set forth above.

The enzymatic conversion of the resulting KDG into 2-deoxy-D-ribose can be achieved by the use of a keto acid decarboxylase. With respect to the preferred embodiments the same applies as has been set forth above.

The enzymatic two step method of converting D-gluconate or D-glucosaminate into 2-deoxy-D-ribose via KDG can be carried out in vitro by using cell extracts of cells expressing the corresponding enzymes or by using purified or partially purified enzymes. The enzymes can be enzymes which are naturally expressed in an organism or they may be recombinantly produced. Methods of preparing and isolating corresponding (recombinant) enzymes are well-known to the person skilled in the art.

In a preferred embodiment the enzymatic two step method of converting D-gluconate or D-glucosaminate into 2-deoxy-D-ribose via KDG is carried out in vivo, i.e. by using a suitable organism, which expresses the required enzyme activities. This organism

may be any type of organism, preferably it is a cell, e.g. a plant, an animal, a fungal cell or a bacterial cell. Most preferably fungal or bacterial cells are used. Preferred fungi are yeasts, such as Saccharomyces cerevisiae; preferred bacterial cells are, e.g. E. coli, Zymomonas mobilis, Zymobacter palmae, Acetobacter pasteurianus, Acinetobacter calcoaceticus, Agrobacterium tumefaciens and Bacillus subtilis. The organism may be an organism which endogenously already expresses one of the enzymatic activities, i.e. a D-gluconate dehydratase or a D-glucosaminate deaminase for producing KDG, or a keto acid decarboxylase for converting KDG into 2-deoxy-D-ribose, and in which the respective other enzymatic activity is expressed due to the introduction of a corresponding exogenous nucleic acid molecule encoding the corresponding enzyme. Alternatively, the organism may also be an organism which naturally does not express the enzyme activities required for converting D-gluconate or D-glucosaminate into KDG and further into 2-deoxy-D-ribose and in which corresponding foreign nucleic acid molecules have been introduced encoding D-gluconate dehydratase or D-glucosaminate deaminase and a keto acid decarboxylase, respectively.

In a particularly, preferred embodiment the organism is an organism which does not express a KDG kinase (kdgK) activity. Such an enzyme activity would lead to a phosphorylation of KDG to KDPG, which in turn is cleaved by an aldolase into pyruvate and glyceraldehyde-phosphate, thereby diverting KDG into a different unwanted metabolic pathway. It is possible to use for the method according to the invention organisms which naturally do not express a kdgK gene. If the used organism naturally expresses a kdgK, means and methods are well-known to the skilled person to produce mutants or variants of such an organism in which the corresponding kdgK gene is inactivated.

If the described method according to the invention is carried out in vivo by using an organism which expresses a D-gluconate dehydratase for converting D-gluconate into KDG and a keto acid decarboxylase for converting KDG into 2-deoxy-D-ribose, this has the advantage that one can provide D-gluconate as a substrate in the culture medium used to culture the organism. D-gluconate is taken up by the organism and is converted into 2-deoxy-D-ribose.

In another particularly, preferred embodiment the organism is an organism which does not express a KDG aldolase (encoded by the eda gene in E. coli) activity. Such an enzyme activity would lead to cleavage of KDG into pyruvate and glyceraldehydes, thereby diverting KDG into a different unwanted metabolic pathway. It is possible to use for the method according to the invention organisms which naturally do not express an eda gene. If the used organism expresses an eda gene, means and methods are well-known to the skilled person to produce mutants or variants of such an organism in which the corresponding eda gene is inactivated.

In still another particularly, preferred embodiment the organism is an organism which does not express a 2-deoxy-D-ribose aldolase (encoded by the deoC gene in E. coli) activity. Such an enzyme activity would lead to cleavage of 2-deoxy-D-ribose into acetaldehyde and glyceraldehyde, thereby diverting 2-deoxy-D-ribose into a different unwanted metabolic pathway. It is possible to use for the method according to the invention organisms which naturally do not express a deoC gene. If the used organism expresses a deoC gene, means and methods are well-known to the skilled person to produce mutants or variants of such an organism in which the corresponding deoC gene is inactivated. For instance a deoC mutant of E. coli has been reported (Valentin-Hansen, EMBO J. 1 (1982), 317-322) as well as a method for deleting the deo operon in E. coli (Kaminski, J. Biol. Chem. 277 (2002), 14400-14407; Valentin-Hansen, Molec. Gen. Genet. 159 (1978), 191-202).

The present invention also relates to organism which are capable of enzymatically converting D-gluconate into KDG due to the expression of a D-gluconate dehydratase and/or of enzymatically converting D-glucosaminate into KDG due to the expression of a D-glucosaminate deaminase and which are furthermore capable of enzymatically converting KDG into 2-deoxy-D-ribose by a decarboxylation reaction catalysed by a keto acid decarboxylase. The organism may in principle be any suitable organism, preferably, it is a cell, e.g. a plant cell, an animal cell, a fungal cell or a bacterial cell. More preferably, it is a fungal or a bacterial cell. Preferred fungi are yeasts, e.g. Saccharomyces cerevisiae. Preferred bacteria are Escherichia coli, Zymomonas mobilis, Zymobacter palmae, Acetobacter pasteurianus, Acinetobacter calcoaceticus, Agrobacterium tumefaciens and Bacillus subtilis. In one aspect, the

organism is an organism which already endogenously expresses a D-gluconate dehydratase or a D-glucosaminate deaminase and into which a foreign nucleic acid molecule has been introduced which encodes a keto acid decarboxylase which can catalyse the decarboxylation of KDG to 2-deoxy-D-ribose. With respect to the preferred embodiments of the keto acid decarboxylase the same applies as has been set forth previously.

In another aspect, the organism is an organism which already expresses a keto acid decarboxylase which is capable of converting KDG into 2-deoxy-D-ribose by a decarboxylation reaction but which does not naturally express a D-gluconate dehydratase or a D-glucosaminate deaminase, and into which a foreign nucleic acid molecule has been introduced which encodes a D-gluconate dehydratase and/or which encodes a D-glucosaminate deaminase. I.e. the organism can be genetically modified so as to express a D-gluconate dehydratase or a D-glucosaminate deaminase or both enzymes.

In a further aspect, the organism is an organism, which naturally does not express a D-gluconate dehydratase, a D-glucosaminate deaminase and a keto acid decarboxylase which is capable of converting KDG by decarboxylation into 2-deoxy-D-ribose, and into which foreign nucleic acid molecules have been introduced encoding a D-gluconate dehydratase or a D-glucosaminate deaminase, or both, and a nucleic acid molecule which encodes a keto acid decarboxylase which is capable of converting KDG into 2-deoxy-D-ribose by decarboxylation.

With respect to the preferred embodiments of the D-gluconate dehydratase, the D-glucosamine deaminase and the keto acid decarboxylase to be expressed in the organisms according to the invention, the same applies which has been set forth above in connection with the method according to the invention.

In a particularly preferred embodiment the organism according to the invention does not express a KDG kinase (kdgK) activity. It can either be an organism which naturally does not express kdgK or it can be an organism which naturally expresses a kdgK but in which the corresponding gene has been inactivated, e.g. by gene disruption or other suitable methods well-known to the person skilled in the art.

The present invention also relates to the use of an enzyme having keto acid decarboxylase activity or of a polynucleotide encoding such an enzyme in a method for converting KDG into 2-deoxy-D-ribose. With respect to the preferred embodiments the same applies as has already been set forth in connection with the method according to the present invention.

These and other embodiments are disclosed and encompassed by the description and examples of the present invention. The disclosure content of any references cited above or below is herewith incorporated into the present application. Further literature concerning any one of the methods, uses and compounds to be employed in accordance with the present invention may be retrieved from public libraries, using for example electronic devices. For example the public database "Medline" may be utilized which is available on the Internet, for example under http://www.ncbi.nlm.nih.gov/PubMed/medline.html. Further databases and addresses, such as http://www.ncbi.nlm.nih.gov/, http://www.infobiogen.fr/, http://www.fmi.ch/biology/research_tools.html, http://www.tigr.org/, are known to the person skilled in the art and can also be obtained using, e.g., http://www.google.de. An overview of patent information in biotechnology and a survey of relevant sources of patent information useful for retrospective searching and for current awareness is given in Berks, TIBTECH 12 (1994), 352-364.
Furthermore, the term "and/or" when occurring herein includes the meaning of "and", "or" and "all or any other combination of the elements connected by said term".

EXAMPLES

Example 1: Cloning of a gene encoding a D-gluconate dehydratase from Agrobacterium tumefaciens strain C58 (CIP 104333)

Agrobacterium tumefaciens strain C58 (CIP 104333) was obtained from Institut Pasteur Collection (CIP, Paris, France). Chromosomal DNA was extracted and a D-gluconate dehydratase gene was amplified by PCR according to standard protocols using the following primers:
5'-CCCTTAATTAATGACGACATCTGATAATCTTC-3', depicted in SEQ ID N° 5;

5'-TTTGCGGCCGCTTAGTGGTTATCGCGCGGC-3', depicted in SEQ ID N° 6;
5'-CCCGGTACCATGACGACATCTGATAATCTTC-3',depicted in SEQ ID N° 7;
A first DNA fragment amplified using the two primers depicted in SEQ ID N° 5 and SEQ ID N° 6, was ligated into a pUC18-derived vector previously digested by PacI and NotI to yield the plasmid pVDM80. A second DNA fragment amplified using the two primers depicted in SEQ ID N° 6 and SEQ ID N° 7, was ligated into a pET29a vector (Novagen) previously digested by KpnI and NotI to yield the plasmid pVDM82. The nucleotide sequence of the cloned gene is depicted in SEQ ID N° 1 and the sequence of the polypeptide encoded by this gene is depicted in SEQ ID N° 2.

Example 2: Expression of a D-gluconate dehydratase activity in Escherichia coli and preparation of 2-dehydro-3-deoxy-D-gluconate from D-gluconate

Competent cells of E. coli BL21 were transformed with the pVDM82 plasmid constructed as described in example 1 yielding strain +1289. Strain + 1289 was cultivated at 30°C in Luria-Bertani (LB) medium (Difco) containing 30 mg/l kanamycin until OD(600 nm) reached a value of 0.6. Then isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a 0.5 mM final concentration. After a further cultivation period of 2 hours and 30 minutes, cells were collected by centrifugation and washed once with 20 mM sodium phosphate buffer pH 7.2. A cell extract was prepared by suspending about 5 g of cells in 10 ml of Tris-HCl 50 mM pH 8.5 buffer containing 10000 units lysozyme (Ready-Lyse, Epicentre, Madison, Wisconsin) and 1 mM EDTA, and incubating the suspension at 30°C for 15 minutes. Then 10000 kUnits deoxyribonuclease I (DNase I, Sigma) as well as 5 mM MgCl2 were added to the preparation which was incubated at 30°C for an additional period of 15 minutes. The cell extract thus obtained was kept frozen at –20°C before use.
1.5 ml of the cell extract was mixed with 2M sodium or potassium D-gluconate in a total volume of 10 ml. This preparation was incubated at 37°C after the pH has been adjusted to 8.5. The progression of 2-dehydro-3-deoxy-D-gluconate (KDG) synthesis was followed by analysing aliquots taken after increasing periods of incubation. Several dilution parts of these aliquots were deposited on silica plates and chromatographied in the following solvent system: isopropanol / water (90/10). A yellow spot of KDG (Rf ~0.40) was detected after revelation with p-anisaldehyde.

KDG was also quantitated using a spectrophotometric assay based on the reaction with semicarbazide hydrochloride as described by Mac Gee (J. Biol. Chem. 1954. 210, 617-626). Typically, after a 30h period of incubation and using the spectrophotometric assay, KDG concentration ranged from 1.5 to 2 M.

The sodium or potassium 2-dehydro-3-deoxy-D-gluconate solution thus obtained could be used as such for further synthetic steps. 2-Dehydro-3-deoxy-D-gluconic acid could also be prepared from such a solution applying published protocols (Bender, Anal. Biochem. 1974. 61, 275-279). A crude preparation of a mixture of 2-dehydro-3-deoxy-D-gluconic acid and KCl could also be obtained by adding one equivalent of HCl to a potassium 2-dehydro-3-deoxy-D-gluconate solution which was then evaporated.

Example 3: Cloning of a gene encoding a D-glucosaminate deaminase from Agrobacterium tumefaciens strain C58 (CIP 104333)

Agrobacterium tumefaciens strain C58 (CIP 104333) was obtained from Institut Pasteur Collection (CIP, Paris, France). Chromosomal DNA was extracted and a D-glucosaminate deaminase gene was amplified by PCR according to standard protocols using the following primers:

5'-CCCTTAATTAATGCAGTCTTCTTCAGCTCTTC-3', depicted in SEQ ID N° 8;

5'-TTTGCGGCCGCCTAGTGAAAGAAGGTTGTGTAGAT-3', depicted in SEQ ID N° 9;

5'-AAATCATGACTATGCAGTCTTCTTCAGCTCTTCG-3', depicted in SEQ ID N° 10;

5'-TATAGATCTCTAGTGAAAGAAGGTTGTGTAGAT-3', depicted in SEQ ID N° 11;

A first DNA fragment amplified using the two primers depicted in SEQ ID N° 8 and SEQ ID N° 9, was ligated into a pUC18-derived vector previously digested by PacI and NotI to yield the plasmid pKDGb1. A second DNA fragment amplified using the two primers depicted in SEQ ID N° 10 and SEQ ID N° 11, was ligated into a pQE60 vector (Qiagen) previously digested by BspH1 and BglII to yield the plasmid pEP18. The nucleotide sequence of the cloned gene is depicted in SEQ ID N° 3 and the sequence of the polypeptide encoded by this gene is depicted in SEQ ID N° 4.

Example 4: Expression of a D-glucosaminate deaminase activity in Escherichia coli and preparation of 2-dehydro-3-deoxy-D-gluconic acid from D-glucosaminate

Competent cells of E. coli MG1655 were transformed with the pEP18 plasmid constructed as described in example 1 and pREP4 (Qiagen) yielding strain +1068. Strain + 1068 was cultivated at 37°C in LB medium containing 30 mg/l kanamycin and 100 mg/l ampicillin until OD(600 nm) reached a value of 0.6. Then IPTG was added to a 0.5 mM final concentration. After a further cultivation period of 2 hours and 30 minutes, cells were collected by centrifugation and washed once with 20 mM sodium phosphate buffer pH 7.2. A cell extract was prepared using the protocol described in example 2.

2 ml of the cell extract was mixed with 100 mM sodium or potassium D-glucosaminate and 0.1 mM pyridoxal phosphate in a total volume of 5 ml. This preparation was incubated at 37°C after the pH has been adjusted to 7.5.
The progression of 2-dehydro-3-deoxy-D-gluconate (KDG) synthesis was followed using the protocols described in example 2. Typically, after a 30h period of incubation and using the spectrophotometric assay described in example 2, KDG concentration ranged from 50 to 100 mM.

Example 5: Preparation of 2-deoxy-D-ribonate from 2-dehydro-3-deoxy-D-gluconate

0.5 ml of a 31% hydrogen peroxyde solution were added to 5 ml of a 1M potassium 2-dehydro-3-deoxy-D-gluconate (KDG) solution at 25°C. The progression of KDG decarboxylation was followed both by the observation of bubbles resulting from the release of carbon dioxide and by the disappearance of KDG using the thin layer chromatography protocol described in example 2. Typically, after a 3h period of reaction the concentration of residual KDG was less than 10 mM.

Example 6: Preparation of 2-deoxy-D-ribitol from 2-deoxy-D-ribonolactone

0.2 g of Rhodium (5 % on carbon) catalyst was added to an aqueous solution of 1 g 2-deoxy-D-ribonolactone prepared following a method described by Deriaz (J. Chem. Soc. (1949), 1879-1883) for the synthesis of 2-deoxy-L-ribonolactone. Hydrogenation of 2-deoxy-D-ribonolactone was performed at 130°C under a pressure of 80 bars. The solution obtained after filtration of the reaction mixture was evaporated. The residue was dissolved in ethyl acetate and further purified by chromatography on a silica column. The solvent was removed in vacuo leading to a yellow oil (yield 85%).The compound thus obtained was identical with 2-deoxy-D-ribitol obtained by reduction of 2-deoxy-D-ribose as described by Rabow (J. Am. Chem. Soc. 122 (1999), 3196-3203).

Example 7: Preparation of 1-N-morpholino-3,4,5-trihydroxypentene-1 from 2-dehydro-3-deoxy-D-gluconate

2 g of 2-dehydro-3-deoxy-D-gluconic acid were suspended in 150 ml benzene. 1.1 ml morpholine and 100 mg p-toluenesulfonic acid were added to the suspension and the reaction mixture was refluxed for 3 hours. Water formed by this reaction was removed by distillation. Benzene was decanted. Solid compounds attached to the vessel were collected, washed with acetone and dried. The main compound present in this preparation (yield 40%) was further purified by column chromatography on a silica column using a gradient of methanol in chloroform. Fractions containing 1-N-morpholino-3,4,5-trihydroxypentene-1 were pooled and solvent was removed in vacuo.

^{1}H-NMR ($D_2$0): δ = 3.15 ppm (4H, t, morpholine), 3.8 ppm (4H, t, morpholine), 3.4 to 4 ppm, (4H, m, 5a-H, 5b-H, 4-H, 3-H), 6.3 and 6.8 ppm (2H, 2d, 1-H and 2-H, J = 4 Hz).

Example 8: Cloning of a gene encoding a pyruvate decarboxylase from Zymomonas mobilis

Zymomonas mobilis strain B-806 (CIP 102538T) was obtained from Institut Pasteur Collection (CIP, Paris, France). Chromosomal DNA was extracted and a pyruvate

decarboxylase gene was amplified by PCR according to standard protocols using the following primers:
5'-GCGTTAATTAATGAGTTATACTGTCGGTACC-3', depicted in SEQ ID N° 12;
5'-TATGCGGCCGCTTAGAGGAGCTTGTTAACAGG-3', depicted in SEQ ID N° 13;
The DNA fragment amplified using the two primers depicted in SEQ ID N° 12 and SEQ ID N° 13, was ligated either into pSP100 or into pEVL5 (respectively a pUC18-derived or a pQE70-derived vector as described below) previously digested by PacI and NotI to yield respectively plasmid pEVL107 and plasmid pEVL420. The nucleotide sequence of the cloned gene as well as the encoded sequence of the corresponding polypeptide can be found at GenBank (accession number AF124349) and is shown in SEQ ID NO: 18 and SEQ ID NO: 19, respectively.

Plasmid pSP100 was obtained by introducing a ribosomal binding site, a PacI and a NotI restriction sites into a pUC18 vector previously digested by EcoRI and BamHI using standard protocols. The complete nucleotide sequence of pSP100 is depicted in SEQ ID N° 14.

Plasmid pEVL5 was obtained by introducing a ribosomal binding site, a PacI and a NotI restriction sites into a pQE70 vector (Qiagen) previously digested by EcoRI and BamHI using standard protocols. The complete nucleotide sequence of pEVL5 is depicted in SEQ ID N° 15.

Example 9: Cloning of a gene encoding a pyruvate decarboxylase from Saccharomyces cerevisiae

Chromosomal DNA was extracted from Saccharomyces cerevisiae strain S288C (ATCC 204508) and a pyruvate decarboxylase gene was amplified by PCR according to standard protocols using the following primers:
5'-ATATTTAATTAATGTCTGAAATTACTTTGG-3', depicted in SEQ ID N° 16;
5'-ATATGCGGCCGCTTATTGCTTAGCGTTGGT-3', depicted in SEQ ID N° 17;
The DNA fragment amplified using the two primers depicted in SEQ ID N° 16 and SEQ ID N° 17, was ligated either into pSP100 or into pEVL5 (respectively a pUC18-derived or a pQE70-derived vector as described in example 8) previously digested by

PacI and NotI to yield respectively plasmid pVDM61 and plasmid pEVL419. The nucleotide sequence of the cloned gene as well as the encoded sequence of the corresponding polypeptide can be found at GenBank (accession number NC001144) and is shown in SEQ ID NO: 20 and SEQ ID NO: 21, respectively.

Example 10: Expression of a pyruvate decarboxylase activity in Escherichia coli and enzymatic synthesis of 2-deoxy-D-ribose from 2-dehydro-3-deoxy-D-gluconate

Expression of pyruvate decarboxylase and preparation of cell-free extracts

Competent cells of E. coli MG1655 strain were transformed with either pEVL107 or pVDM61 (constructed as described in examples 8 and 9) yielding respectively strain +1735 and strain + 844. These strains were cultivated at 37°C in Luria-Bertani (LB) medium (Difco) containing 100 mg/l ampicillin until OD(600 nm) reached a value around 1.5.

Competent cells of E. coli MG1655 strain harbouring pREP4 plasmid (Qiagen) were transformed with either pEVL420 or pEVL419 (constructed as described in Examples 8 and 9) yielding respectively strain +3150 and +3148. These strains were cultivated at 37°C in Luria-Bertani (LB) medium (Difco) containing 100 mg/l ampicillin and 30 mg/l kanamycin until OD(600 nm) reached a value of 0.6. Then isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a 0.5 mM final concentration. After a further cultivation period of 2 hours and 30 minutes, cells were collected by centrifugation and washed once with 20 mM sodium phosphate buffer pH 7.2.

For each strain a cell-free extract was prepared using the same protocol as described in Example 2. Then crude cell-free extracts were passed through a PD-10 column (Amersham) equilibrated with 50 mM Tris-acetate buffer pH 6 and stored at –20°C.

Enzymatic synthesis of 2-deoxy-D-ribose from 2-dehydro-3-deoxy-D-gluconate

1.0 ml of cell-free extract was mixed with 20 mM sodium 2-dehydro-3-deoxy-D-gluconate, 0.5 mM thiamine pyrophosphate and 5 mM $MgCl_2$ in a total volume of 1.5 ml of 50 mM tris-acetate buffer pH 6. The progression of 2-deoxy-D-ribose (DRI) synthesis was followed by analysing aliquots taken after increasing periods of

incubation at 37°C. About 1 µl of each aliquot which had been previously concentrated five-fold by evaporation was deposited on a silica plate and chromatographied in the following solvent system: butanol / triethylamine / water (10/2/5). A blue spot of DRI (Rf ~0.50) was detected after revelation with orcinol when using cell-free extracts of either strain +3150 or +3148 after a period of incubation of 65 hours. The crude preparation containing the spot corresponding to DRI was concentrated and passed through a 1.5 ml silica column equilibrated with isopropanol. The fractions containing the expected DRI compound were pooled, concentrated and the resulting sample analysed by mass spectrometry. The results of such an analysis confirmed the identity of the isolated compound with DRI, and the production of DRI from KDG catalysed by pyruvate decarboxylase either from Zymomonas mobilis or from Saccharomyces cerevisiae.

Example 11: Cloning of a gene encoding a pyruvate decarboxylase from Acetobacter pasteurianus, expression of encoded pyruvate decarboxylase activity in Escherichia coli and enzymatic synthesis of 2-deoxy-D-ribose from 2-dehydro-3-deoxy-D-gluconate

Acetobacter pasteurianus strain NCIB 8618 (DSMZ 2347) was obtained from DSMZ Collection (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany). Chromosomal DNA was extracted from the cells and a pyruvate decarboxylase gene was amplified by PCR according to standard protocols using the following primers:

5'-TCTTTAATTAATGGGTTGTCCGTCATTCATATA-3', depicted in SEQ ID N° 22;

5'-CTAAAGCTTTTAGGCCAGAGTGGTCTTGCGCG-3', depicted in SEQ ID N° 23;

The DNA fragment amplified using the two primers depicted in SEQ ID N° 22 and SEQ ID N° 23, was ligated either into pSP100 or into pEVL5 (respectively a pUC18-derived or a pQE70-derived vector as described in example 8) previously digested by PacI and NotI to yield respectively plasmid pEVL541 and plasmid pEVL560. The nucleotide sequence SEQ ID N° 24 of the cloned gene as well as the encoded sequence of the corresponding polypeptide SEQ ID N° 25 can be found at GenBank (accession number AF368435).

Competent cells of E. coli MG1655 strain were transformed with pEVL541 yielding strain +3559. Competent cells of E. coli MG1655 strain harbouring pREP4 plasmid (Qiagen) were transformed with pEVL560 yielding strain +3924. These strains were cultivated and cell-free extracts were prepared as described in Example 10. Cell-free extracts were incubated with KDG and the progression of 2-deoxy-D-ribose (DRI) synthesis was followed as described in Example 10. A spot corresponding to DRI was observed indicating that pyruvate decarboxylase from Acetobacter pasteurianus was able to decarboxylate KDG into DRI.

Example 12: Cloning of a gene encoding a pyruvate decarboxylase from Zymobacter palmae, expression of encoded pyruvate decarboxylase activity in Escherichia coli and enzymatic synthesis of 2-deoxy-D-ribose from 2-dehydro-3-deoxy-D-gluconate

Zymobacter palmae strain T109 (DSMZ10491) was obtained from DSMZ Collection (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany). Chromosomal DNA was extracted from the cells and a pyruvate decarboxylase gene was amplified by PCR according to standard protocols using the following primers:

5'-ATCTTAATTAATGTATACCGTTGGTATGTACT-3', depicted in SEQ ID N° 26;

5'-TATGCGGCCGCTTACGCTTGTGGTTTGCGAGAGT-3', depicted in SEQ ID N° 27;

The DNA fragment amplified using the two primers depicted in SEQ ID N° 26 and SEQ ID N° 27, was ligated either into pSP100 or into pEVL5 (respectively a pUC18-derived or a pQE70-derived vector as described in example 8) previously digested by PacI and NotI to yield respectively plasmid pEVL546 and plasmid pEVL561. The nucleotide sequence of the cloned gene as well as the encoded sequence of the corresponding polypeptide is shown in SEQ ID NOs: 28 and 29, respectively and can be found at GenBank (accession number AF474145).

Competent cells of E. coli MG1655 strain were transformed with pEVL546 yielding strain +3568. Competent cells of E. coli MG1655 strain harbouring pREP4 plasmid (Qiagen) were transformed with pEVL560 yielding strain +3923. These strains were cultivated and cell-free extracts were prepared as described in Example 10. Cell-free

extracts were incubated with KDG and the progression of 2-deoxy-D-ribose (DRI) synthesis was followed as described in Example 10. A spot corresponding to DRI was observed indicating that pyruvate decarboxylase from Zymobacter palmae was able to decarboxylate KDG into DRI.

Example 13. Cloning of a gene encoding a benzoylformate decarboxylase from Pseudomonas putida, expression of encoded benzoylformate decarboxylase activity in Escherichia coli and enzymatic synthesis of 2-deoxy-D-ribose from 2-dehydro-3-deoxy-D-gluconate.

Pseudomonas putida strain Migula (DSMZ 291) was obtained from DSMZ Collection (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany). Chromosomal DNA was extracted from and a benzoylformate decarboxylase gene was amplified by PCR according to standard protocols using the following primers:

5'-CTATTAATTAATGGCTTCGGTACACGGCACCA-3', depicted in SEQ ID N° 30;

5'-TATGCGGCCGCTTACTTCACCGGGCTTACGGTGC-3', depicted in SEQ ID N° 31;

The DNA fragment amplified using the two primers depicted in SEQ ID N° 30 and SEQ ID N° 31, was ligated either into pSP100 or into pEVL5 (respectively a pUC18-derived or a pQE70-derived vector as described in example 8) previously digested by PacI and NotI to yield respectively plasmid pEVL681 and plasmid pEVL670. The nucleotide sequence SEQ ID N° 32 of the cloned gene as well as the encoded sequence of the corresponding polypeptide SEQ ID N° 33 can be found at GenBank (accessing number AY143338).

Competent cells of E. coli MG1655 strain were transformed with pEVL681 yielding strain +4050. Competent cells of E. coli MG1655 strain harbouring pREP4 plasmid (Qiagen) were transformed with pEVL670 yielding strain +3927. Those strains were cultivated and cell-free extracts were prepared as described in example 10. Cell-free extracts were incubated with KDG and the progression of 2-deoxy-D-ribose (DRI) synthesis was followed as described in example 10. A spot corresponding to DRI was observed indicating that benzoylformate decarboxylase from *Pseudomonas putida* was able to decarboxylate KDG into DRI.

Preparative enzymatic synthesis of 2-deoxy-D-ribose

100 µl of cell-free extract from strain +3927 (containing 2.5 mg of bacterial proteins) were mixed with 300 mM sodium 2-dehydro-3-deoxy-D-gluconate, 0.5 mM thiamine pyrophosphate and 5 mM $MgCl_2$ in a total volume of 0.5 ml of 80 mM potassium phosphate buffer pH 6. After a period of incubation of 16 and 40 hours, few µl of a solution of HCl 2N were added to the incubation mixture until the pH reached a value of 6. The progression of 2-deoxy-D-ribose (DRI) synthesis was also followed by analysing aliquots taken after increasing periods of incubation at 37°C. About 1 µl of each aliquot was deposited on a silica plate and chromatographied as described in example 10. The concentration of 2-deoxy-D-ribose was estimated to be about 200 mM by comparison with standard solutions. 13C NMR analysis of the crude mixture confirmed that the compound formed from 2-dehydro-3-deoxy-D-gluconate was 2-deoxy-D-ribose, and that the concentration of 2-deoxy-D-ribose was closed to 25 g/l. Another preparative enzymatic synthesis was performed in the same conditions except that no addition of acid was made along the incubation period. In those conditions, the concentration of 2-deoxy-D-ribose was closed to 10 g/l, far lower than the concentration reached in the preceding experiment for which the pH had been controlled and regularly adjusted to a value of 6.

Example 14. Enzymatic synthesis of 2-deoxy-D-ribose from D-gluconate

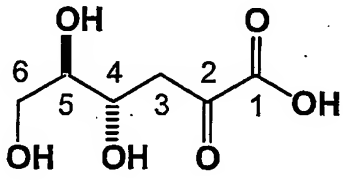
One pot enzymatic synthesis of 2-deoxy-D-ribose from D-gluconate was achieved as follows, using D-gluconate dehydratase encoded by gcnD gene of Agrobacterium tumefaciens and pyruvate decarboxylase from Zymomonas mobilis:

50 µl of cell-free extract from strain +1289 (containing 1.5 mg of bacterial proteins) and 400 µl of cell-free extract from strain +3150 (containing 17 mg of bacterial proteins after concentration by ultrafiltration) prepared as described respectively in example 2 and in example 10, were mixed with 50 mM potassium D-gluconate, 0.5 mM thiamine pyrophosphate and 5 mM $MgCl_2$ in a total volume of 0.5 ml of 50 mM N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES) buffer pH 7. The progression of 2-deoxy-D-ribose (DRI) synthesis was also followed by analysing

aliquots taken after increasing periods of incubation at 37°C. After a period of incubation of 18 hours, about 1 µl of the incubation mixture was deposited on a silica plate and chromatographied as described in example 10. The concentration of 2-deoxy-D-ribose was estimated to be about 1 g/l by comparison with standard solutions.

Claims

1. A method for producing 2'-deoxynucleosides or 2'-deoxynucleoside precursors from a compound of formula (I) or its salts



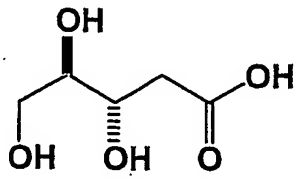
(I)

or a protected form thereof in a process comprising a decarboxylation step.

2. The method of claim 1 wherein the decarboxylation step cleaves the C1-C2 bond of the compound of formula (I) or its salts or a protected form thereof.

3. The method of claim 1 or 2, wherein the decarboxylation step is directly carried out on the compound of formula (I) or its salts or a protected form thereof.

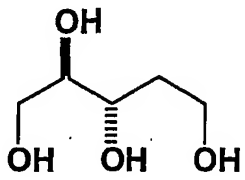
4. The method of any of claims 1 to 3, wherein the decarboxylation step takes place by reacting the compound of formula (I) or its salts or a protected form thereof with hydrogen peroxide to yield a compound of formula (II) or its salts.



(II)

or a protected form thereof as a 2'-deoxynucleoside precursor.

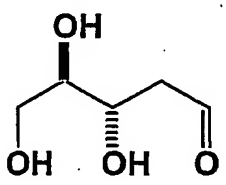
5. The method of claim 4, further comprising the conversion of the compound of formula (II) or its salts or a protected form thereof into a compound of formula (IV)



(IV)

or a protected form thereof as a 2'-deoxynucleoside precursor.

6. The method of claim 4, further comprising the conversion of the compound of formula (II) or its salts or a protected form thereof into a compound of formula (III)

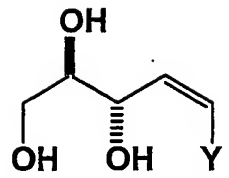


(III)

or a protected form thereof as a 2'-deoxynucleoside precursor.

7. The method of claim 6, comprising the conversion of the compound of formula (II) or its salts or a protected form thereof into the compound of formula (IV) or a protected form thereof as an intermediate which is then converted to the compound of formula (III) or a protected form thereof.

8. The method of any of claims 1 to 3, wherein the decarboxylation step takes place by reacting the compound of formula (I) or its salts or a protected form thereof with an amine Y-H, wherein H represents a hydrogen atom bound to the nitrogen atom of the amino group, to produce a compound of formula (V),



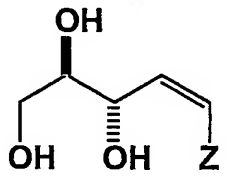
(V)

or its respective trans isomer or a protected form thereof, as a 2'-deoxynucleoside precursor.

9. The method of claim 8, wherein Y-H represents a linear or cyclic secondary amine.

10. The method of claims 8 or 9, wherein Y-H is morpholine, pyrrolidine, piperidine, N-methyl piperazine or diethylamine.

11. The method of any of claims 8 to 10, further comprising the step of reacting a compound of formula (V) or its trans isomer or a protected form thereof with Z-H, wherein H represents a hydrogen atom and Z represents a leaving group, to produce a compound of formula (VI)

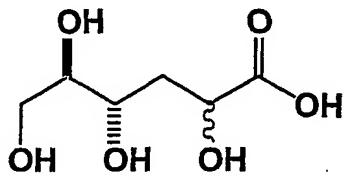


(VI)

or its respective trans isomer or a protected form thereof, as a 2'-deoxynucleoside precursor.

12. The method of claim 11, wherein Z-H is water, to produce a compound of formula (III) or a protected form thereof as a 2'-deoxynucleoside precursor.

13. The method of claim 1 or 2, wherein the compound of formula (I) or its salts or a protected form thereof is converted to a compound of formula (VII), or its salts or a protected form thereof or a mixture of the respective epimers,



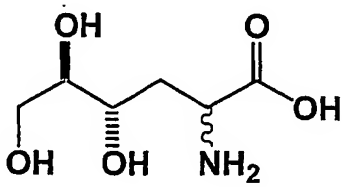
(VII)

which is then decarboxylated to yield a compound of formula (III) or a protected form thereof as a 2'-deoxynucleoside precursor.

14. The method of claim 13, wherein the conversion of (I) or its salts or a protected form thereof to (VII) or a protected form thereof takes place by reduction with sodium borohydride or by hydrogenation using Nickel Raney or Platinum oxide catalyst.

15. The method of claim 13 to 14, wherein the decarboxylation step takes place by reaction with hydrogen peroxide.

16. The method of claim 1 or 2, wherein the compound of formula (I) or its salts or a protected form thereof is converted to a compound of formula (VIII), or its salts or a protected form thereof or a mixture of the respective epimers,



(VIII)

which is then decarboxylated to yield a compound of formula (III) or a protected form thereof as a 2'-deoxynucleoside precursor.

17. The method of claim 16, wherein a compound of formula (VIII) or a protected form thereof or a mixture of the respective epimers is reacted with ninhydrin, thereby leading to the compound (III) or a protected form thereof.

18. The method of claim 16 or 17, wherein the conversion of (I) or its salts or a protected form thereof to (VIII) or a protected form thereof takes place by reductive amination with ammonia and sodium cyanoborohydride.

19. The method of any of claims 1 to 18, wherein the protective group(s) are independently chosen from acetate ester, benzoate ester, allyl ether, benzyl ether, trityl ether, ter-butyldimethylsilyl (TBDMS) ether, isopropylidene or a benzylidene acetal.

20. The method of any one of claims 1 to 3, wherein the decarboxylation step is effected by an enzymatic reaction comprising a single step.

21. The method of claim 20, wherein the enzymatic reaction is catalysed by an enzyme having keto acid decarboxylase activity.

22. The method of claim 21, wherein the enzyme having keto acid decarboxylase activity is a thiamine pyrophosphate (TPP) dependent keto acid decarboxylase.

23. The method of claim 22, wherein the TPP dependent keto acid decarboxylase is a pyruvate decarboxylase (EC 4.1.1.1), a benzoylformate decarboxylase (EC 4.1.1.7), an indolepyruvate decarboxylase (EC 4.1.1.74), a phosphonopyruvate decarboxylase, a sulfopyruvate decarboxylase (EC 4.1.1.79), an oxalyl-coenzyme A decarboxylase (EC 4.1.1.8), an oxoglutarate decarboxylase (EC 4.1.1.71) or a phenylpyruvate decarboxylase (EC 4.1.1.43).

24. The method of claim 23, wherein the pyruvate decarboxylase is of eukaryotic origin.

25. The method of claim 24, wherein the eukaryotic organism is a yeast organism.

26. The method of claim 25, wherein the yeast is Saccharomyces cerevisiae.

27. The method of claim 23, wherein the pyruvate decarboxylase is of prokaryotic origin.

28. The method of claim 27, wherein the prokaryotic organism is of the genus Zymomonas, Zymobacter or Acetobacter.

29. The method of claim 28, wherein the organism is of the species Zymomonas mobilis, Zymobacter plamae or Acetobacter pasteurianus.

30. The method of claim 23, wherein the benzoylformate decarboxylase is of prokaryotic origin.

31. The method of claim 30, wherein the prokaryotic organism is of the genus Pseudomonas.

32. The method of claim 31, wherein the organism is of the species Pseudomonas putida.

33. The method of any one of the claims 20 to 32, wherein the pH is regulated by addition of an acid between pH 5 and pH 9.

34. The method of claim 33, wherein the pH value is regulated between pH 6 and pH 8.

35. The method of claim 33 or 34, wherein the acid is HCl, H_2SO_4, D-gluconic acid or 2-dehydro-3-deoxy-D-gluconic acid.

36. The method of any one of claims 1 to 35, comprising the preliminary step of producing the compound of formula (I) from D-gluconate or a D-gluconate salt by the use of a gluconate dehydratase activity.

37. The method of claim 36, wherein the D-gluconate salt is potassium or sodium D-gluconate.

38. The method of claims 36 or 37, wherein the gluconate dehydratase is encoded by a polynucleotide comprising the nucleotide sequence selected from the group consisting of:

(a) nucleotide sequences encoding a polypeptide comprising the amino acid sequence of SEQ ID N°2;

(b) nucleotide sequences comprising the coding sequence of SEQ ID N°1;

(c) nucleotide sequences encoding a fragment encoded by a nucleotide sequence of (a) or (b);

(d) nucleotide sequences hybridising with a nucleotide sequence of any one of (a) to (c); and

(e) nucleotide sequences which deviate from the nucleoside sequence of (d) as a result of degeneracy of the genetic code.

39. The method of any one of claims 1 to 35, comprising the preliminary step of producing the compound of formula (I) from D-glucosaminate by the use of a glucosaminate deaminase activity.

40. The method of claim 39, wherein the glucosaminate deaminase is encoded by a polynucleotide comprising the nucleotide sequence selected from the group consisting of:

(a) nucleotide sequences encoding a polypeptide comprising the amino acid sequence of SEQ ID N°4;

(b) nucleotide sequences comprising the coding sequence of SEQ ID N°3;

(c) nucleotide sequences encoding a fragment encoded by a nucleotide sequence of (a) or (b);

(d) nucleotide sequences hybridising with a nucleotide sequence of any one of (a) to (c); and

(e) nucleotide sequences which deviate from the nucleoside sequence of (d) as a result of degeneracy of the genetic code.

41. An organism which is capable of enzymatically converting D-gluconate into 2-dehydro-3-deoxy-D-gluconatedue to the expression of a D-gluconate dehydratase and/or capable of enzymatically converting D-glucosaminate into 2-dehydro-3-deoxy-D-gluconatedue to the expression of a D-glucosaminate deaminase and which is capable of enzymatically converting 2-dehydro-3-deoxy-D-gluconateby decarboxylation into 2-deoxy-D-ribose due to the expression of a keto acid decarboxylase.

42. The organism of claim 41 which does not express a 2-dehydro-3-deoxy-D-gluconatekinase activity.

43. The organism of claim 41 or 42 which does not express a 2-dehydro-3-deoxy-D-gluconatealdolase activity.

44. The organism of any one of claims 41 to 43 which does not express a 2-deoxy-D-ribose aldolase activity.

45. The method of any of claims 20 to 40 which is carried out by using an organism according to any one of claims 41 to 44.

46 Use of a polynucleotide as defined in claim 38 or of a gluconate dehydratase encoded by such a polynucleotide in a method according to claims 36 or 37.

47. Use of a polynucleotide as defined in claim 40 or of a glucosaminate deaminase encoded by such a polynucleotide in a method according to claim 39.

48. Use of an enzyme having keto acid decarboxylase activity or of a polynucleotide encoding such an enzyme in a method for converting a compound of the formula (I) into 2-deoxy-D-ribose.

SEQUENCE LISTING

<110> Evologic S.A.

Maliere Technologies Société Civile

Rhodia Chimie

Marliere, Phillipe

<120> Cloning of gluconate dehydratase gcnD gene

<130> G 3111 EP

<160> 33

<170> PatentIn version 3.1

<210> 1

<211> 1812

<212> DNA

<213> Agrobacterium tumefaciens

<220>

<221> CDS

<222> (1)..(1809)

<223>

<400> 1

```
atg acg aca tct gat aat ctt cct gca act cag ggc aag ctc cgt tcg       48
Met Thr Thr Ser Asp Asn Leu Pro Ala Thr Gln Gly Lys Leu Arg Ser
1               5                   10                  15

cgc gcc tgg ttc gac aac cca gcc aat gcg gac atg acc gcg ctt tat       96
Arg Ala Trp Phe Asp Asn Pro Ala Asn Ala Asp Met Thr Ala Leu Tyr
            20                  25                  30

ctc gag cgt tac atg aac ttc ggt ctc agc cag gcc gag ctt cag tcc      144
Leu Glu Arg Tyr Met Asn Phe Gly Leu Ser Gln Ala Glu Leu Gln Ser
```

```
        35                  40                  45
gac cgc ccg att atc ggt att gcg cag acc ggt tcc gac ctt tcg ccc      192
Asp Arg Pro Ile Ile Gly Ile Ala Gln Thr Gly Ser Asp Leu Ser Pro
    50                  55                  60

tgc aac cgt cat cat ctg gaa ctc gcc aac cgt ctg cgt gaa ggc att      240
Cys Asn Arg His His Leu Glu Leu Ala Asn Arg Leu Arg Glu Gly Ile
65                  70                  75                  80

cgt gaa gcc ggc ggc atc gcc atc gaa ttc ccg gtg cat ccg atc cag      288
Arg Glu Ala Gly Gly Ile Ala Ile Glu Phe Pro Val His Pro Ile Gln
                85                  90                  95

gaa acc ggt aag cgt ccg aca gcg ggc ctt gat cgc aac ctg gct tac      336
Glu Thr Gly Lys Arg Pro Thr Ala Gly Leu Asp Arg Asn Leu Ala Tyr
            100                 105                 110

ctc ggc ctc gtg gaa gtg ctt tat ggc tat ccg ctc gac ggc gtt gtt      384
Leu Gly Leu Val Glu Val Leu Tyr Gly Tyr Pro Leu Asp Gly Val Val
        115                 120                 125

ctg acc atc ggc tgc gac aag acc acg cct gcc tgt ctt atg gcg gcg      432
Leu Thr Ile Gly Cys Asp Lys Thr Thr Pro Ala Cys Leu Met Ala Ala
    130                 135                 140

gcc acc gtc aac att ccg gcc atc gcc ctt tcc gtc ggt ccc atg ctg      480
Ala Thr Val Asn Ile Pro Ala Ile Ala Leu Ser Val Gly Pro Met Leu
145                 150                 155                 160

aac ggc tgg ttc cgc ggt gag cgc acc ggt tcc ggc acc atc gtc tgg      528
Asn Gly Trp Phe Arg Gly Glu Arg Thr Gly Ser Gly Thr Ile Val Trp
                165                 170                 175

aag gcc cgc gaa ctg ctg gcg aag ggc gag atc gat tac cag ggc ttc      576
Lys Ala Arg Glu Leu Leu Ala Lys Gly Glu Ile Asp Tyr Gln Gly Phe
            180                 185                 190

gtc aag ctc gtt gcc tcg tct gcc ccg tcc acc ggc tat tgc aac acc      624
Val Lys Leu Val Ala Ser Ser Ala Pro Ser Thr Gly Tyr Cys Asn Thr
        195                 200                 205

atg ggc acg gca aca acc atg aac tcg ctc gcc gaa gcg ctc ggc atg      672
Met Gly Thr Ala Thr Thr Met Asn Ser Leu Ala Glu Ala Leu Gly Met
    210                 215                 220

cag ctt ccc ggc tcc gcc gcc att ccg gcg cct tac cgt gac cgt cag      720
Gln Leu Pro Gly Ser Ala Ala Ile Pro Ala Pro Tyr Arg Asp Arg Gln
225                 230                 235                 240

gaa gtc tct tac ctc acc ggc ctg cgc atc gtc gac atg gtc agg gaa      768
Glu Val Ser Tyr Leu Thr Gly Leu Arg Ile Val Asp Met Val Arg Glu
                245                 250                 255

gac ctg aaa cca tca gac atc atg acc aag gat gcc ttc atc aac gcc      816
Asp Leu Lys Pro Ser Asp Ile Met Thr Lys Asp Ala Phe Ile Asn Ala
            260                 265                 270

atc cgc gtt aat tcg gcg atc ggc ggt tcc acc aac gcg ccg atc cat      864
Ile Arg Val Asn Ser Ala Ile Gly Gly Ser Thr Asn Ala Pro Ile His
```

```
      275                 280                 285

cta aac ggc ctt gcc cgc cat gtc ggc gtc gag ctg acg gtg gat gac      912
Leu Asn Gly Leu Ala Arg His Val Gly Val Glu Leu Thr Val Asp Asp
    290                 295                 300

tgg cag acc tat ggc gaa gac gtg ccg ctg ctc gtc aac ctg cag ccg      960
Trp Gln Thr Tyr Gly Glu Asp Val Pro Leu Leu Val Asn Leu Gln Pro
305                 310                 315                 320

gca ggc gaa tat ctc ggc gag gac tat tac cat gcc ggc ggc gtt ccc     1008
Ala Gly Glu Tyr Leu Gly Glu Asp Tyr Tyr His Ala Gly Gly Val Pro
                325                 330                 335

gct gtc gtc aat cag ctg atg acc caa ggg ctg atc atg gaa gac gcc     1056
Ala Val Val Asn Gln Leu Met Thr Gln Gly Leu Ile Met Glu Asp Ala
            340                 345                 350

atg acc gtc aac ggc aag acc atc ggc gac aat tgc cgt ggc gcg atc     1104
Met Thr Val Asn Gly Lys Thr Ile Gly Asp Asn Cys Arg Gly Ala Ile
        355                 360                 365

atc gaa gac gag aag gtc atc cgc ccc tat gag cag ccg ctc aag gag     1152
Ile Glu Asp Glu Lys Val Ile Arg Pro Tyr Glu Gln Pro Leu Lys Glu
    370                 375                 380

cgt gcc ggc ttc cgc gtt ctg cgc ggc aat ctg ttc tcc tcg gcc atc     1200
Arg Ala Gly Phe Arg Val Leu Arg Gly Asn Leu Phe Ser Ser Ala Ile
385                 390                 395                 400

atg aag aca agc gtg att tcg gaa gaa ttc cgc ggt cgt tac ctc tcc     1248
Met Lys Thr Ser Val Ile Ser Glu Glu Phe Arg Gly Arg Tyr Leu Ser
                405                 410                 415

aac cct gat gat ccg gaa gcc ttc gaa ggc cgc gcc gtg gtg ttc gat     1296
Asn Pro Asp Asp Pro Glu Ala Phe Glu Gly Arg Ala Val Val Phe Asp
            420                 425                 430

ggt ccg gag gat tac cat cat cgc atc gac gat ccg tcg ctt ggc atc     1344
Gly Pro Glu Asp Tyr His His Arg Ile Asp Asp Pro Ser Leu Gly Ile
        435                 440                 445

gac gcc aac acc gtc ctg ttc atg cgc ggc gcc ggt ccg atc ggt tac     1392
Asp Ala Asn Thr Val Leu Phe Met Arg Gly Ala Gly Pro Ile Gly Tyr
    450                 455                 460

ccg ggc gca gcg gaa gtg gtg aac atg cgc gcg ccg gat tac ctt ctg     1440
Pro Gly Ala Ala Glu Val Val Asn Met Arg Ala Pro Asp Tyr Leu Leu
465                 470                 475                 480

aag caa ggc gtc agt tcg ctg ccc tgc atc ggc gat ggc cgc cag tcc     1488
Lys Gln Gly Val Ser Ser Leu Pro Cys Ile Gly Asp Gly Arg Gln Ser
                485                 490                 495

ggc acg tcg ggc agc cca tcc atc ctc aat gcc tcg ccg gaa gcg gcg     1536
Gly Thr Ser Gly Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu Ala Ala
            500                 505                 510

gcc ggc ggc ggt ctg tct att ctg cag acg ggt gac cgc gtc cgc atc     1584
```

```
Ala Gly Gly Gly Leu Ser Ile Leu Gln Thr Gly Asp Arg Val Arg Ile
        515                 520                 525

gat gtg ggc cgc ggc aag gcc gat atc ctg ata tca ggt gaa gag ctc    1632
Asp Val Gly Arg Gly Lys Ala Asp Ile Leu Ile Ser Gly Glu Glu Leu
    530                 535                 540

gcc aag cgt tac gag gcg ctg gca gct cag ggc ggt tat aag ttc ccc    1680
Ala Lys Arg Tyr Glu Ala Leu Ala Ala Gln Gly Gly Tyr Lys Phe Pro
545                 550                 555                 560

gac cac cag acg ccg tgg cag gaa atc cag cgc ggt atc gtc agc cag    1728
Asp His Gln Thr Pro Trp Gln Glu Ile Gln Arg Gly Ile Val Ser Gln
                565                 570                 575

atg gaa acc ggc gcg gtt ctg gaa ccg gcc gta aag tat cag cgc atc    1776
Met Glu Thr Gly Ala Val Leu Glu Pro Ala Val Lys Tyr Gln Arg Ile
            580                 585                 590

gcc cag acc aag ggc ctg ccg cgc gat aac cac tga                   1812
Ala Gln Thr Lys Gly Leu Pro Arg Asp Asn His
        595                 600
```

<210> 2

<211> 603

<212> PRT

<213> Agrobacterium tumefaciens

<400> 2

```
Met Thr Thr Ser Asp Asn Leu Pro Ala Thr Gln Gly Lys Leu Arg Ser
1               5                   10                  15

Arg Ala Trp Phe Asp Asn Pro Ala Asn Ala Asp Met Thr Ala Leu Tyr
            20                  25                  30

Leu Glu Arg Tyr Met Asn Phe Gly Leu Ser Gln Ala Glu Leu Gln Ser
        35                  40                  45

Asp Arg Pro Ile Ile Gly Ile Ala Gln Thr Gly Ser Asp Leu Ser Pro
    50                  55                  60

Cys Asn Arg His His Leu Glu Leu Ala Asn Arg Leu Arg Glu Gly Ile
65                  70                  75                  80

Arg Glu Ala Gly Gly Ile Ala Ile Glu Phe Pro Val His Pro Ile Gln
                85                  90                  95
```

```
Glu Thr Gly Lys Arg Pro Thr Ala Gly Leu Asp Arg Asn Leu Ala Tyr
            100                 105                 110

Leu Gly Leu Val Glu Val Leu Tyr Gly Tyr Pro Leu Asp Gly Val Val
        115                 120                 125

Leu Thr Ile Gly Cys Asp Lys Thr Thr Pro Ala Cys Leu Met Ala Ala
    130                 135                 140

Ala Thr Val Asn Ile Pro Ala Ile Ala Leu Ser Val Gly Pro Met Leu
145                 150                 155                 160

Asn Gly Trp Phe Arg Gly Glu Arg Thr Gly Ser Gly Thr Ile Val Trp
                165                 170                 175

Lys Ala Arg Glu Leu Leu Ala Lys Gly Glu Ile Asp Tyr Gln Gly Phe
            180                 185                 190

Val Lys Leu Val Ala Ser Ser Ala Pro Ser Thr Gly Tyr Cys Asn Thr
        195                 200                 205

Met Gly Thr Ala Thr Thr Met Asn Ser Leu Ala Glu Ala Leu Gly Met
    210                 215                 220

Gln Leu Pro Gly Ser Ala Ala Ile Pro Ala Pro Tyr Arg Asp Arg Gln
225                 230                 235                 240

Glu Val Ser Tyr Leu Thr Gly Leu Arg Ile Val Asp Met Val Arg Glu
                245                 250                 255

Asp Leu Lys Pro Ser Asp Ile Met Thr Lys Asp Ala Phe Ile Asn Ala
            260                 265                 270

Ile Arg Val Asn Ser Ala Ile Gly Gly Ser Thr Asn Ala Pro Ile His
        275                 280                 285

Leu Asn Gly Leu Ala Arg His Val Gly Val Glu Leu Thr Val Asp Asp
    290                 295                 300

Trp Gln Thr Tyr Gly Glu Asp Val Pro Leu Leu Val Asn Leu Gln Pro
305                 310                 315                 320

Ala Gly Glu Tyr Leu Gly Glu Asp Tyr Tyr His Ala Gly Gly Val Pro
                325                 330                 335
```

```
Ala Val Val Asn Gln Leu Met Thr Gln Gly Leu Ile Met Glu Asp Ala
            340                 345                 350

Met Thr Val Asn Gly Lys Thr Ile Gly Asp Asn Cys Arg Gly Ala Ile
        355                 360                 365

Ile Glu Asp Glu Lys Val Ile Arg Pro Tyr Glu Gln Pro Leu Lys Glu
    370                 375                 380

Arg Ala Gly Phe Arg Val Leu Arg Gly Asn Leu Phe Ser Ser Ala Ile
385                 390                 395                 400

Met Lys Thr Ser Val Ile Ser Glu Glu Phe Arg Gly Arg Tyr Leu Ser
                405                 410                 415

Asn Pro Asp Asp Pro Glu Ala Phe Glu Gly Arg Ala Val Val Phe Asp
            420                 425                 430

Gly Pro Glu Asp Tyr His His Arg Ile Asp Asp Pro Ser Leu Gly Ile
        435                 440                 445

Asp Ala Asn Thr Val Leu Phe Met Arg Gly Ala Gly Pro Ile Gly Tyr
    450                 455                 460

Pro Gly Ala Ala Glu Val Val Asn Met Arg Ala Pro Asp Tyr Leu Leu
465                 470                 475                 480

Lys Gln Gly Val Ser Ser Leu Pro Cys Ile Gly Asp Gly Arg Gln Ser
                485                 490                 495

Gly Thr Ser Gly Ser Pro Ser Ile Leu Asn Ala Ser Pro Glu Ala Ala
            500                 505                 510

Ala Gly Gly Gly Leu Ser Ile Leu Gln Thr Gly Asp Arg Val Arg Ile
        515                 520                 525

Asp Val Gly Arg Gly Lys Ala Asp Ile Leu Ile Ser Gly Glu Glu Leu
    530                 535                 540

Ala Lys Arg Tyr Glu Ala Leu Ala Ala Gln Gly Gly Tyr Lys Phe Pro
545                 550                 555                 560

Asp His Gln Thr Pro Trp Gln Glu Ile Gln Arg Gly Ile Val Ser Gln
```

565 570 575

Met Glu Thr Gly Ala Val Leu Glu Pro Ala Val Lys Tyr Gln Arg Ile
580 585 590

Ala Gln Thr Lys Gly Leu Pro Arg Asp Asn His
595 600

<210> 3

<211> 1272

<212> DNA

<213> Agrobacterium tumefaciens

<220>

<221> CDS

<222> (1)..(1269)

<223>

<400> 3

```
atg cag tct tct tca gct ctt cgg caa tca acc ggc gat cag tcg gaa       48
Met Gln Ser Ser Ser Ala Leu Arg Gln Ser Thr Gly Asp Gln Ser Glu
1               5                   10                  15

tac cat gcc cag tcg aat atg atc ggc tct agc ccg gcg gac ggt ttg       96
Tyr His Ala Gln Ser Asn Met Ile Gly Ser Ser Pro Ala Asp Gly Leu
            20                  25                  30

ctc gca ttg ccg ctt ctg acc gtc gat ctt gcc gtc tat cgc ggt aat      144
Leu Ala Leu Pro Leu Leu Thr Val Asp Leu Ala Val Tyr Arg Gly Asn
        35                  40                  45

cgg gat cgc ttt ctt gcg ctt gtc tcg gcc cat gga gcg aag gcg gct      192
Arg Asp Arg Phe Leu Ala Leu Val Ser Ala His Gly Ala Lys Ala Ala
    50                  55                  60

cca cat gcc aag acg ccg atg tgc ccg gag atc gcg atc gat ctg att      240
Pro His Ala Lys Thr Pro Met Cys Pro Glu Ile Ala Ile Asp Leu Ile
65                  70                  75                  80

gaa gcc ggt gcc tgg ggc gcg acg gtc gcc gat ctc ttc cag gcg gaa      288
Glu Ala Gly Ala Trp Gly Ala Thr Val Ala Asp Leu Phe Gln Ala Glu
                85                  90                  95

gtc ctg ctc aag gcc ggc gtg tcg aac ata ttg atc gcc aac cag atc      336
Val Leu Leu Lys Ala Gly Val Ser Asn Ile Leu Ile Ala Asn Gln Ile
            100                 105                 110
```

```
ggc gga ttg aca tcc gcc aga cgc cta cgc atg ctc gca gat gct ttt      384
Gly Gly Leu Thr Ser Ala Arg Arg Leu Arg Met Leu Ala Asp Ala Phe
        115                 120                 125

ccg aaa gcc gag att atc tgc tgt gtc gat tct gtt cag gcc tcg gcc      432
Pro Lys Ala Glu Ile Ile Cys Cys Val Asp Ser Val Gln Ala Ser Ala
    130                 135                 140

aat ctg gtt cag gcc ttt caa ggg cgt gtg gat gcc cca ttc aag gtc      480
Asn Leu Val Gln Ala Phe Gln Gly Arg Val Asp Ala Pro Phe Lys Val
145                 150                 155                 160

ttc atc gaa gtc ggt gtc ggc cgc act ggc gcc cgt acg ttg aat gtt      528
Phe Ile Glu Val Gly Val Gly Arg Thr Gly Ala Arg Thr Leu Asn Val
                165                 170                 175

gca aag gat atc atc gac acc atc tcg aca agt gca gaa atc gta ctg      576
Ala Lys Asp Ile Ile Asp Thr Ile Ser Thr Ser Ala Glu Ile Val Leu
            180                 185                 190

gcc ggt gtg tcg acc tat gaa ggc tcc gtc tcc ggg gaa acg tcg gaa      624
Ala Gly Val Ser Thr Tyr Glu Gly Ser Val Ser Gly Glu Thr Ser Glu
        195                 200                 205

gca ctc gat gca aac atg gcg gcc ctg ttc gat ctc ctg acc gac agt      672
Ala Leu Asp Ala Asn Met Ala Ala Leu Phe Asp Leu Leu Thr Asp Ser
    210                 215                 220

ctt gca tcg ata cgc gaa aaa gat ccc ggg cgc ccg cta acg gtt tca      720
Leu Ala Ser Ile Arg Glu Lys Asp Pro Gly Arg Pro Leu Thr Val Ser
225                 230                 235                 240

gcc ggc ggt tcg atc cat ttc gac cgc gtg ctc gcg gcg ctt gtg ccc      768
Ala Gly Gly Ser Ile His Phe Asp Arg Val Leu Ala Ala Leu Val Pro
                245                 250                 255

gtt tgc gag gcg gat ggc aat gcg acg ttg ttg ctg cgc agc ggc gcc      816
Val Cys Glu Ala Asp Gly Asn Ala Thr Leu Leu Leu Arg Ser Gly Ala
            260                 265                 270

atc ttc ttc tct gat cac ggt gta tat cag cgc ggt ttc cag gca gtc      864
Ile Phe Phe Ser Asp His Gly Val Tyr Gln Arg Gly Phe Gln Ala Val
        275                 280                 285

gac gcc cgc aac cta ctc gca tcc ggc aag gtt gtc ttc aag gca tcc      912
Asp Ala Arg Asn Leu Leu Ala Ser Gly Lys Val Val Phe Lys Ala Ser
    290                 295                 300

gag gca ttt cag ccc tca atg cga atc tgg gcg gag gtc atc tcc gtt      960
Glu Ala Phe Gln Pro Ser Met Arg Ile Trp Ala Glu Val Ile Ser Val
305                 310                 315                 320

cct gag ccg ggg ctg gcg atc gtc ggc atg ggc atg cgg gat gta tcg     1008
Pro Glu Pro Gly Leu Ala Ile Val Gly Met Gly Met Arg Asp Val Ser
                325                 330                 335

ttc gat cag gac ctg ccc gtg gcg ctt cgg ctc cat agg gac gga cat     1056
Phe Asp Gln Asp Leu Pro Val Ala Leu Arg Leu His Arg Asp Gly His
```

```
            340                 345                 350

ctg gtc gaa gct gat ctc tct tca tcc gcg aag gtc ggc aag ctc aat     1104
Leu Val Glu Ala Asp Leu Ser Ser Ser Ala Lys Val Gly Lys Leu Asn
        355                 360                 365

gac cag cat gcc ttc ttg tcc ttc ggg aac ggc agc agt ctg gca atc     1152
Asp Gln His Ala Phe Leu Ser Phe Gly Asn Gly Ser Ser Leu Ala Ile
    370                 375                 380

ggc gat gtc ata gaa ttc ggc atc tcg cat ccc tgc acg tgc ttc gat     1200
Gly Asp Val Ile Glu Phe Gly Ile Ser His Pro Cys Thr Cys Phe Asp
385                 390                 395                 400

cgc tgg cgc gtc ttt cac gga atc gat gga tca ggc cgg atc cag cgc     1248
Arg Trp Arg Val Phe His Gly Ile Asp Gly Ser Gly Arg Ile Gln Arg
                405                 410                 415

atc tac aca acc ttc ttt cac tag                                    1272
Ile Tyr Thr Thr Phe Phe His
            420
```

<210> 4

<211> 423

<212> PRT

<213> Agrobacterium tumefaciens

<400> 4

```
Met Gln Ser Ser Ser Ala Leu Arg Gln Ser Thr Gly Asp Gln Ser Glu
1               5                   10                  15

Tyr His Ala Gln Ser Asn Met Ile Gly Ser Ser Pro Ala Asp Gly Leu
            20                  25                  30

Leu Ala Leu Pro Leu Leu Thr Val Asp Leu Ala Val Tyr Arg Gly Asn
        35                  40                  45

Arg Asp Arg Phe Leu Ala Leu Val Ser Ala His Gly Ala Lys Ala Ala
    50                  55                  60

Pro His Ala Lys Thr Pro Met Cys Pro Glu Ile Ala Ile Asp Leu Ile
65                  70                  75                  80

Glu Ala Gly Ala Trp Gly Ala Thr Val Ala Asp Leu Phe Gln Ala Glu
                85                  90                  95
```

```
Val Leu Leu Lys Ala Gly Val Ser Asn Ile Leu Ile Ala Asn Gln Ile
            100                 105                 110

Gly Gly Leu Thr Ser Ala Arg Arg Leu Arg Met Leu Ala Asp Ala Phe
        115                 120                 125

Pro Lys Ala Glu Ile Ile Cys Cys Val Asp Ser Val Gln Ala Ser Ala
    130                 135                 140

Asn Leu Val Gln Ala Phe Gln Gly Arg Val Asp Ala Pro Phe Lys Val
145                 150                 155                 160

Phe Ile Glu Val Gly Val Gly Arg Thr Gly Ala Arg Thr Leu Asn Val
                165                 170                 175

Ala Lys Asp Ile Ile Asp Thr Ile Ser Thr Ser Ala Glu Ile Val Leu
            180                 185                 190

Ala Gly Val Ser Thr Tyr Glu Gly Ser Val Ser Gly Glu Thr Ser Glu
        195                 200                 205

Ala Leu Asp Ala Asn Met Ala Ala Leu Phe Asp Leu Leu Thr Asp Ser
    210                 215                 220

Leu Ala Ser Ile Arg Glu Lys Asp Pro Gly Arg Pro Leu Thr Val Ser
225                 230                 235                 240

Ala Gly Gly Ser Ile His Phe Asp Arg Val Leu Ala Ala Leu Val Pro
                245                 250                 255

Val Cys Glu Ala Asp Gly Asn Ala Thr Leu Leu Leu Arg Ser Gly Ala
            260                 265                 270

Ile Phe Phe Ser Asp His Gly Val Tyr Gln Arg Gly Phe Gln Ala Val
        275                 280                 285

Asp Ala Arg Asn Leu Leu Ala Ser Gly Lys Val Val Phe Lys Ala Ser
    290                 295                 300

Glu Ala Phe Gln Pro Ser Met Arg Ile Trp Ala Glu Val Ile Ser Val
305                 310                 315                 320

Pro Glu Pro Gly Leu Ala Ile Val Gly Met Gly Met Arg Asp Val Ser
                325                 330                 335
```

Phe Asp Gln Asp Leu Pro Val Ala Leu Arg Leu His Arg Asp Gly His
340 345 350

Leu Val Glu Ala Asp Leu Ser Ser Ser Ala Lys Val Gly Lys Leu Asn
355 360 365

Asp Gln His Ala Phe Leu Ser Phe Gly Asn Gly Ser Ser Leu Ala Ile
370 375 380

Gly Asp Val Ile Glu Phe Gly Ile Ser His Pro Cys Thr Cys Phe Asp
385 390 395 400

Arg Trp Arg Val Phe His Gly Ile Asp Gly Ser Gly Arg Ile Gln Arg
405 410 415

Ile Tyr Thr Thr Phe Phe His
420

<210> 5

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> oligonucleotide primer

<400> 5
cccttaatta atgacgacat ctgataatct tc 32

<210> 6

<211> 30

<212> DNA

<213> artificial sequence

<220>

<223> oligonucleotide primer

<400> 6
tttgcggccg cttagtggtt atcgcgcggc 30

<210> 7

<211> 31

<212> DNA

<213> artificial sequence

<220>

<223> oligonucleotide primer

<400> 7
cccggtacca tgacgacatc tgataatctt c 31

<210> 8

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> oligonucleotide primer

<400> 8
cccttaatta atgcagtctt cttcagctct tc 32

<210> 9

<211> 35

<212> DNA

<213> artificial sequence

<220>

<223> oligonucleotide primer

<400> 9
tttgcggccg cctagtgaaa gaaggttgtg tagat 35

<210> 10

<211> 34

<212> DNA

<213> artificial sequence

<220>

<223> oligonucleotide primer

<400> 10
aaatcatgac tatgcagtct tcttcagctc ttcg 34

<210> 11

<211> 33

<212> DNA

<213> artificial sequence

<220>

<223> oligonucleotide primer

<400> 11
tatagatctc tagtgaaaga aggttgtgta gat 33

<210> 12

<211> 31

<212> DNA

<213> artificial sequence

<220>

<223> artificial sequence

<400> 12
gcgttaatta atgagttata ctgtcggtac c 31

<210> 13

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> artificial sequence

<400> 13
tatgcggccg cttagaggag cttgttaaca gg 32

<210> 14

<211> 2665

<212> DNA

<213> vector

<400> 14
gacgaaaggg cctcgtgata cgcctatttt ataggttaa tgtcatgata ataatggttt 60

cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aacccctatt tgtttatttt 120

tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat 180

aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt 240

ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 300

ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 360

tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc 420

tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggt cgccgcatac 480

actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 540

gcatgacagt aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca 600

acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg 660

gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg 720

acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg 780

gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag 840

ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 900

gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 960

cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac 1020

agatcgctga gataggtgcc tcactgatta agcattggta actgtcagac caagtttact 1080

catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga 1140

tccttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt 1200

cagaccccgt agaaaagatc aaaggatctt cttgagatcc ttttttctg cgcgtaatct 1260

gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc 1320

taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 1380

ttctagtgta gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc 1440

tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 1500

ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggtt 1560

cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg 1620

agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg 1680

gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt 1740

atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag 1800

gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt 1860

gctggccttt tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta 1920

ttaccgcctt tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt 1980

cagtgagcga ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc 2040

cgattcatta atgcagctgg cacgacaggt ttcccgactg gaaagcgggc agtgagcgca 2100

acgcaattaa tgtgagttag ctcactcatt aggcacccca ggctttacac tttatgcttc 2160

cggctcgtat gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg 2220

accatgatta cgaattcgag gttaattaac cccgcggccg caagcttggc actggccgtc 2280

gttttacaac gtcgtgactg ggaaaaccct ggcgttaccc aacttaatcg ccttgcagca 2340

catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa 2400

cagttgcgca gcctgaatgg cgaatggcgc ctgatgcggt attttctcct tacgcatctg 2460

tgcggtattt cacaccgcat atggtgcact ctcagtacaa tctgctctga tgccgcatag 2520

ttaagccagc cccgacaccc gccaacaccc gctgacgcgc cctgacgggc ttgtctgctc 2580

ccggcatccg cttacagaca agctgtgacc gtctccggga gctgcatgtg tcagaggttt 2640

tcaccgtcat caccgaaacg cgcga 2665

<210> 15

<211> 3433

<212> DNA

<213> vector

<400> 15

ctcgagaaat cataaaaaat ttatttgctt tgtgagcgga taacaattat aatagattca 60

attgtgagcg gataacaatt tcacacagaa ttcttaaaga ggagaaatta attaaccccg 120

cggccgcgga tccagatctc atcaccatca ccatcactaa gcttaattag ctgagcttgg 180

actcctgttg atagatccag taatgacctc agaactccat ctggatttgt tcagaacgct 240

cggttgccgc cgggcgtttt ttattggtga gaatccaagc tagcttggcg agattttcag 300

gagctaagga agctaaaatg gagaaaaaaa tcactggata taccaccgtt gatatatccc 360

aatggcatcg taaagaacat tttgaggcat ttcagtcagt tgctcaatgt acctataacc 420

agaccgttca gctggatatt acggcctttt taaagaccgt aaagaaaaat aagcacaagt 480

tttatccggc ctttattcac attcttgccc gcctgatgaa tgctcatccg gaatttcgta 540

tggcaatgaa agacggtgag ctggtgatat gggatagtgt tcacccttgt tacaccgttt 600

tccatgagca aactgaaacg ttttcatcgc tctggagtga ataccacgac gatttccggc 660

agtttctaca catatattcg caagatgtgg cgtgttacgg tgaaaacctg gcctatttcc 720

ctaaagggtt tattgagaat atgtttttcg tctcagccaa tccctgggtg agtttcacca 780

gttttgattt aaacgtggcc aatatggaca acttcttcgc ccccgttttc accatgggca 840

aatattatac gcaaggcgac aaggtgctga tgccgctggc gattcaggtt catcatgccg 900

tttgtgatgg cttccatgtc ggcagaatgc ttaatgaatt acaacagtac tgcgatgagt 960

ggcagggcgg ggcgtaattt ttttaaggca gttattggtg cccttaaacg cctggggtaa 1020

tgactctcta gcttgaggca tcaaataaaa cgaaaggctc agtcgaaaga ctgggccttt 1080

cgttttatct gttgtttgtc ggtgaacgct ctcctgagta ggacaaatcc gccctctaga 1140

gctgcctcgc gcgtttcggt gatgacggtg aaaacctctg acacatgcag ctcccggaga 1200

cggtcacagc ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag ggcgcgtcag 1260

cgggtgttgg cgggtgtcgg ggcgcagcca tgacccagtc acgtagcgat agcggagtgt 1320

atactggctt aactatgcgg catcagagca gattgtactg agagtgcacc atatgcggtg 1380

tgaaataccg cacagatgcg taaggagaaa ataccgcatc aggcgctctt ccgcttcctc 1440

gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa 1500

ggcggtaata cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa 1560

aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt tccataggct 1620

ccgccccccct gacgagcatc acaaaaatcg acgctcaagt cagaggtggc gaaacccgac 1680

aggactataa agataccagg cgtttccccc tggaagctcc ctcgtgcgct ctcctgttcc 1740

gaccctgccg cttaccggat acctgtccgc ctttctccct tcgggaagcg tggcgctttc 1800

tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca agctgggctg 1860

tgtgcacgaa ccccccgttc agcccgaccg ctgcgccttа tccggtaact atcgtcttga 1920

gtccaacccg gtaagacacg acttatcgcc actggcagca gccactggta acaggattag 1980

cagagcgagg tatgtaggcg gtgctacaga gttcttgaag tggtggccta actacggcta 2040

cactagaagg acagtatttg gtatctgcgc tctgctgaag ccagttacct tcggaaaaag 2100

agttggtagc tcttgatccg gcaaacaaac caccgctggt agcggtggtt tttttgtttg 2160

caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga tcttttctac 2220

ggggtctgac gctcagtgga acgaaaactc acgttaaggg attttggtca tgagattatc 2280

aaaaaggatc ttcacctaga tccttttaaa ttaaaaatga agttttaaat caatctaaag 2340

tatatatgag taaacttggt ctgacagtta ccaatgctta atcagtgagg cacctatctc 2400

agcgatctgt ctatttcgtt catccatagt tgcctgactc cccgtcgtgt agataactac 2460

gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgcgag acccacgctc 2520

accggctcca gatttatcag caataaacca gccagccgga agggccgagc gcagaagtgg 2580

tcctgcaact ttatccgcct ccatccagtc tattaattgt tgccgggaag ctagagtaag 2640

tagttcgcca gttaatagtt tgcgcaacgt tgttgccatt gctacaggca tcgtggtgtc 2700

acgctcgtcg tttggtatgg cttcattcag ctccggttcc caacgatcaa ggcgagttac 2760

atgatccccc atgttgtgca aaaaagcggt tagctccttc ggtcctccga tcgttgtcag 2820

aagtaagttg gccgcagtgt tatcactcat ggttatggca gcactgcata attctcttac 2880

tgtcatgcca tccgtaagat gcttttctgt gactggtgag tactcaacca agtcattctg 2940

agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg ataataccgc 3000

gccacatagc agaactttaa aagtgctcat cattggaaaa cgttcttcgg ggcgaaaact 3060

ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg cacccaactg 3120

atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa 3180

tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac tcttcctttt 3240

tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca tatttgaatg 3300

tatttagaaa aataaacaaa taggggttcc gcgcacattt ccccgaaaag tgccacctga 3360

cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta tcacgaggcc 3420

ctttcgtctt cac 3433

<210> 16

<211> 30

<212> DNA

<213> artificial sequence

<220>

<223> artificial sequence

<400> 16
atatttaatt aatgtctgaa attactttgg 30

<210> 17

<211> 30

<212> DNA

<213> artificial sequence

<220>

<223> artificial sequence

<400> 17
atatgcggcc gcttattgct tagcgttggt 30

<210> 18

<211> 1707

<212> DNA

<213> Zymomonas mobilis

<220>

<221> CDS

<222> (1)..(1707)

<223>

<400> 18

```
atg agt tat act gtc ggt acc tat tta gcg gag cgg ctt gtc cag att       48
Met Ser Tyr Thr Val Gly Thr Tyr Leu Ala Glu Arg Leu Val Gln Ile
1               5                   10                  15

ggt ctc aag cat cac ttc gca gtc gcg ggc gac tac aac ctc gtc ctt       96
Gly Leu Lys His His Phe Ala Val Ala Gly Asp Tyr Asn Leu Val Leu
            20                  25                  30

ctt gac aac ctg ctt ttg aac aaa aac atg gag cag gtt tat tgc tgt      144
Leu Asp Asn Leu Leu Leu Asn Lys Asn Met Glu Gln Val Tyr Cys Cys
        35                  40                  45

aac gaa ctg aac tgc ggt ttc agt gca gaa ggt tat gct cgt gcc aaa      192
Asn Glu Leu Asn Cys Gly Phe Ser Ala Glu Gly Tyr Ala Arg Ala Lys
    50                  55                  60

ggc gca gca gca gcc gtc gtt acc tac agc gtc ggt gcg ctt tcc gca      240
Gly Ala Ala Ala Ala Val Val Thr Tyr Ser Val Gly Ala Leu Ser Ala
65                  70                  75                  80

ttt gat gct atc ggt ggc gcc tat gca gaa aac ctt ccg gtt atc ctg      288
Phe Asp Ala Ile Gly Gly Ala Tyr Ala Glu Asn Leu Pro Val Ile Leu
                85                  90                  95

atc tcc ggt gct ccg aac aac aat gat cac gct gct ggt cac gtg ttg      336
Ile Ser Gly Ala Pro Asn Asn Asn Asp His Ala Ala Gly His Val Leu
            100                 105                 110

cat cac gct ctt ggc aaa acc gac tat cac tat cag ttg gaa atg gcc      384
His His Ala Leu Gly Lys Thr Asp Tyr His Tyr Gln Leu Glu Met Ala
        115                 120                 125

aag aac atc acg gcc gcc gct gaa gcg att tac acc ccg gaa gaa gct      432
Lys Asn Ile Thr Ala Ala Ala Glu Ala Ile Tyr Thr Pro Glu Glu Ala
    130                 135                 140

ccg gct aaa atc gat cac gtg att aaa act gct ctt cgt gag aag aag      480
Pro Ala Lys Ile Asp His Val Ile Lys Thr Ala Leu Arg Glu Lys Lys
145                 150                 155                 160

ccg gtt tat ctc gaa atc gct tgc aac att gct tcc atg ccc tgc gcc      528
Pro Val Tyr Leu Glu Ile Ala Cys Asn Ile Ala Ser Met Pro Cys Ala
                165                 170                 175

gct cct gga ccg gca agc gca ttg ttc aat gac gaa gcc agc gac gaa      576
Ala Pro Gly Pro Ala Ser Ala Leu Phe Asn Asp Glu Ala Ser Asp Glu
            180                 185                 190

gct tct ttg aat gca gcg gtt gaa gaa acc ctg aaa ttc atc gcc aac      624
Ala Ser Leu Asn Ala Ala Val Glu Glu Thr Leu Lys Phe Ile Ala Asn
        195                 200                 205

cgc gac aaa gtt gcc gtc ctc gtc ggc agc aag ctg cgc gca gct ggt      672
Arg Asp Lys Val Ala Val Leu Val Gly Ser Lys Leu Arg Ala Ala Gly
    210                 215                 220

gct gaa gaa gct gct gtc aaa ttt gct gat gct ctc ggt ggc gca gtt      720
```

```
Ala Glu Glu Ala Ala Val Lys Phe Ala Asp Ala Leu Gly Gly Ala Val
225                 230                 235                 240

gct acc atg gct gct gca aaa agc ttc ttc cca gaa gaa aac ccg cat      768
Ala Thr Met Ala Ala Ala Lys Ser Phe Phe Pro Glu Glu Asn Pro His
                245                 250                 255

tac atc ggc acc tca tgg ggt gaa gtc agc tat ccg ggc gtt gaa aag      816
Tyr Ile Gly Thr Ser Trp Gly Glu Val Ser Tyr Pro Gly Val Glu Lys
            260                 265                 270

acg atg aaa gaa gcc gat gcg gtt atc gct ctg gct cct gtc ttc aac      864
Thr Met Lys Glu Ala Asp Ala Val Ile Ala Leu Ala Pro Val Phe Asn
        275                 280                 285

gac tac tcc acc act ggt tgg acg gat att cct gat cct aag aaa ctg      912
Asp Tyr Ser Thr Thr Gly Trp Thr Asp Ile Pro Asp Pro Lys Lys Leu
    290                 295                 300

gtt ctc gct gaa ccg cgt tct gtc gtc gtt aac ggc att cgc ttc ccc      960
Val Leu Ala Glu Pro Arg Ser Val Val Val Asn Gly Ile Arg Phe Pro
305                 310                 315                 320

agc gtc cat ctg aaa gac tat ctg acc cgt ttg gct cag aaa gtt tcc     1008
Ser Val His Leu Lys Asp Tyr Leu Thr Arg Leu Ala Gln Lys Val Ser
                325                 330                 335

aag aaa acc ggt gca ttg gac ttc ttc aaa tcc ctc aat gca ggt gaa     1056
Lys Lys Thr Gly Ala Leu Asp Phe Phe Lys Ser Leu Asn Ala Gly Glu
            340                 345                 350

ctg aag aaa gcc gct ccg gct gat ccg agt gct ccg ttg gtc aac gca     1104
Leu Lys Lys Ala Ala Pro Ala Asp Pro Ser Ala Pro Leu Val Asn Ala
        355                 360                 365

gaa atc gcc cgt cag gtc gaa gct ctt ctg acc ccg aac acg acg gtt     1152
Glu Ile Ala Arg Gln Val Glu Ala Leu Leu Thr Pro Asn Thr Thr Val
    370                 375                 380

att gct gaa acc ggt gac tct tgg ttc aat gct cag cgc atg aag ctc     1200
Ile Ala Glu Thr Gly Asp Ser Trp Phe Asn Ala Gln Arg Met Lys Leu
385                 390                 395                 400

ccg aac ggt gct cgc gtt gaa tat gaa atg cag tgg ggt cac att ggt     1248
Pro Asn Gly Ala Arg Val Glu Tyr Glu Met Gln Trp Gly His Ile Gly
                405                 410                 415

tgg tcc gtt cct gcc gcc ttc ggt tat gcc gtc ggt gct ccg gaa cgt     1296
Trp Ser Val Pro Ala Ala Phe Gly Tyr Ala Val Gly Ala Pro Glu Arg
            420                 425                 430

cgc aac atc ctc atg gtt ggt gat ggt tcc ttc cag ctg acg gct cag     1344
Arg Asn Ile Leu Met Val Gly Asp Gly Ser Phe Gln Leu Thr Ala Gln
        435                 440                 445

gaa gtc gct cag atg gtt cgc ctg aaa ctg ccg gtt atc atc ttc ttg     1392
Glu Val Ala Gln Met Val Arg Leu Lys Leu Pro Val Ile Ile Phe Leu
    450                 455                 460
```

```
atc aat aac tat ggt tac acc atc gaa gtt atg atc cat gat ggt ccg     1440
Ile Asn Asn Tyr Gly Tyr Thr Ile Glu Val Met Ile His Asp Gly Pro
465                 470                 475                 480

tac aac aac atc aag aac tgg gat tat gcc ggt ctg atg gaa gtg ttc     1488
Tyr Asn Asn Ile Lys Asn Trp Asp Tyr Ala Gly Leu Met Glu Val Phe
                485                 490                 495

aac ggt aac ggt ggt tat gac agc ggt gct ggt aaa ggc ctg aag gct     1536
Asn Gly Asn Gly Gly Tyr Asp Ser Gly Ala Gly Lys Gly Leu Lys Ala
            500                 505                 510

aaa acc ggt ggc gaa ctg gca gaa gct atc aag gtt gct ctg gca aac     1584
Lys Thr Gly Gly Glu Leu Ala Glu Ala Ile Lys Val Ala Leu Ala Asn
        515                 520                 525

acc gac ggc cca acc ctg atc gaa tgc ttc atc ggt cgt gaa gac tgc     1632
Thr Asp Gly Pro Thr Leu Ile Glu Cys Phe Ile Gly Arg Glu Asp Cys
    530                 535                 540

act gaa gaa ttg gtc aaa tgg ggt aag cgc gtt gct gcc gcc aac agc     1680
Thr Glu Glu Leu Val Lys Trp Gly Lys Arg Val Ala Ala Ala Asn Ser
545                 550                 555                 560

cgt aag cct gtt aac aag ctc ctc tag                               1707
Arg Lys Pro Val Asn Lys Leu Leu
                565
```

<210> 19

<211> 568

<212> PRT

<213> Zymomonas mobilis

<400> 19

```
Met Ser Tyr Thr Val Gly Thr Tyr Leu Ala Glu Arg Leu Val Gln Ile
1               5                   10                  15

Gly Leu Lys His His Phe Ala Val Ala Gly Asp Tyr Asn Leu Val Leu
            20                  25                  30

Leu Asp Asn Leu Leu Leu Asn Lys Asn Met Glu Gln Val Tyr Cys Cys
        35                  40                  45

Asn Glu Leu Asn Cys Gly Phe Ser Ala Glu Gly Tyr Ala Arg Ala Lys
    50                  55                  60

Gly Ala Ala Ala Ala Val Val Thr Tyr Ser Val Gly Ala Leu Ser Ala
65                  70                  75                  80
```

```
Phe Asp Ala Ile Gly Gly Ala Tyr Ala Glu Asn Leu Pro Val Ile Leu
                85                  90                  95

Ile Ser Gly Ala Pro Asn Asn Asn Asp His Ala Ala Gly His Val Leu
           100                 105                 110

His His Ala Leu Gly Lys Thr Asp Tyr His Tyr Gln Leu Glu Met Ala
       115                 120                 125

Lys Asn Ile Thr Ala Ala Ala Glu Ala Ile Tyr Thr Pro Glu Glu Ala
   130                 135                 140

Pro Ala Lys Ile Asp His Val Ile Lys Thr Ala Leu Arg Glu Lys Lys
145                 150                 155                 160

Pro Val Tyr Leu Glu Ile Ala Cys Asn Ile Ala Ser Met Pro Cys Ala
                165                 170                 175

Ala Pro Gly Pro Ala Ser Ala Leu Phe Asn Asp Glu Ala Ser Asp Glu
           180                 185                 190

Ala Ser Leu Asn Ala Ala Val Glu Glu Thr Leu Lys Phe Ile Ala Asn
       195                 200                 205

Arg Asp Lys Val Ala Val Leu Val Gly Ser Lys Leu Arg Ala Ala Gly
   210                 215                 220

Ala Glu Glu Ala Ala Val Lys Phe Ala Asp Ala Leu Gly Gly Ala Val
225                 230                 235                 240

Ala Thr Met Ala Ala Ala Lys Ser Phe Phe Pro Glu Glu Asn Pro His
                245                 250                 255

Tyr Ile Gly Thr Ser Trp Gly Glu Val Ser Tyr Pro Gly Val Glu Lys
           260                 265                 270

Thr Met Lys Glu Ala Asp Ala Val Ile Ala Leu Ala Pro Val Phe Asn
       275                 280                 285

Asp Tyr Ser Thr Thr Gly Trp Thr Asp Ile Pro Asp Pro Lys Lys Leu
   290                 295                 300

Val Leu Ala Glu Pro Arg Ser Val Val Val Asn Gly Ile Arg Phe Pro
```

305 310 315 320

Ser Val His Leu Lys Asp Tyr Leu Thr Arg Leu Ala Gln Lys Val Ser
325 330 335

Lys Lys Thr Gly Ala Leu Asp Phe Phe Lys Ser Leu Asn Ala Gly Glu
340 345 350

Leu Lys Lys Ala Ala Pro Ala Asp Pro Ser Ala Pro Leu Val Asn Ala
355 360 365

Glu Ile Ala Arg Gln Val Glu Ala Leu Leu Thr Pro Asn Thr Thr Val
370 375 380

Ile Ala Glu Thr Gly Asp Ser Trp Phe Asn Ala Gln Arg Met Lys Leu
385 390 395 400

Pro Asn Gly Ala Arg Val Glu Tyr Glu Met Gln Trp Gly His Ile Gly
405 410 415

Trp Ser Val Pro Ala Ala Phe Gly Tyr Ala Val Gly Ala Pro Glu Arg
420 425 430

Arg Asn Ile Leu Met Val Gly Asp Gly Ser Phe Gln Leu Thr Ala Gln
435 440 445

Glu Val Ala Gln Met Val Arg Leu Lys Leu Pro Val Ile Ile Phe Leu
450 455 460

Ile Asn Asn Tyr Gly Tyr Thr Ile Glu Val Met Ile His Asp Gly Pro
465 470 475 480

Tyr Asn Asn Ile Lys Asn Trp Asp Tyr Ala Gly Leu Met Glu Val Phe
485 490 495

Asn Gly Asn Gly Gly Tyr Asp Ser Gly Ala Gly Lys Gly Leu Lys Ala
500 505 510

Lys Thr Gly Gly Glu Leu Ala Glu Ala Ile Lys Val Ala Leu Ala Asn
515 520 525

Thr Asp Gly Pro Thr Leu Ile Glu Cys Phe Ile Gly Arg Glu Asp Cys
530 535 540

Thr Glu Glu Leu Val Lys Trp Gly Lys Arg Val Ala Ala Ala Asn Ser
545 550 555 560

Arg Lys Pro Val Asn Lys Leu Leu
565

<210> 20

<211> 1692

<212> DNA

<213> Saccharomyces cerevisiae

<220>

<221> CDS

<222> (1)..(1692)

<223>

<400> 20

```
atg tct gaa att act ttg ggt aaa tat ttg ttc gaa aga tta aag caa       48
Met Ser Glu Ile Thr Leu Gly Lys Tyr Leu Phe Glu Arg Leu Lys Gln
1               5                  10                  15

gtc aac gtt aac acc gtt ttc ggt ttg cca ggt gac ttc aac ttg tcc       96
Val Asn Val Asn Thr Val Phe Gly Leu Pro Gly Asp Phe Asn Leu Ser
            20                  25                  30

ttg ttg gac aag atc tac gaa gtt gaa ggt atg aga tgg gct ggt aac      144
Leu Leu Asp Lys Ile Tyr Glu Val Glu Gly Met Arg Trp Ala Gly Asn
        35                  40                  45

gcc aac gaa ttg aac gct gct tac gcc gct gat ggt tac gct cgt atc      192
Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Ile
    50                  55                  60

aag ggt atg tct tgt atc atc acc acc ttc ggt gtc ggt gaa ttg tct      240
Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser
65                  70                  75                  80

gct ttg aac ggt att gcc ggt tct tac gct gaa cac gtc ggt gtt ttg      288
Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu
                85                  90                  95

cac gtt gtt ggt gtc cca tcc atc tct gct caa gct aag caa ttg ttg      336
His Val Val Gly Val Pro Ser Ile Ser Ala Gln Ala Lys Gln Leu Leu
            100                 105                 110

ttg cac cac acc ttg ggt aac ggt gac ttc act gtt ttc cac aga atg      384
Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met
```

```
        115                 120                 125

tct gcc aac att tct gaa acc act gct atg atc act gac att gct acc      432
Ser Ala Asn Ile Ser Glu Thr Thr Ala Met Ile Thr Asp Ile Ala Thr
    130                 135                 140

gcc cca gct gaa att gac aga tgt atc aga acc act tac gtc acc caa      480
Ala Pro Ala Glu Ile Asp Arg Cys Ile Arg Thr Thr Tyr Val Thr Gln
145                 150                 155                 160

aga cca gtc tac tta ggt ttg cca gct aac ttg gtc gac ttg aac gtc      528
Arg Pro Val Tyr Leu Gly Leu Pro Ala Asn Leu Val Asp Leu Asn Val
                165                 170                 175

cca gct aag ttg ttg caa act cca att gac atg tct ttg aag cca aac      576
Pro Ala Lys Leu Leu Gln Thr Pro Ile Asp Met Ser Leu Lys Pro Asn
            180                 185                 190

gat gct gaa tcc gaa aag gaa gtc att gac acc atc ttg gct ttg gtc      624
Asp Ala Glu Ser Glu Lys Glu Val Ile Asp Thr Ile Leu Ala Leu Val
        195                 200                 205

aag gat gct aag aac cca gtt atc ttg gct gat gct tgt tgt tcc aga      672
Lys Asp Ala Lys Asn Pro Val Ile Leu Ala Asp Ala Cys Cys Ser Arg
    210                 215                 220

cac gac gtc aag gct gaa act aag aag ttg att gac ttg act caa ttc      720
His Asp Val Lys Ala Glu Thr Lys Lys Leu Ile Asp Leu Thr Gln Phe
225                 230                 235                 240

cca gct ttc gtc acc cca atg ggt aag ggt tcc att gac gaa caa cac      768
Pro Ala Phe Val Thr Pro Met Gly Lys Gly Ser Ile Asp Glu Gln His
                245                 250                 255

cca aga tac ggt ggt gtt tac gtc ggt acc ttg tcc aag cca gaa gtt      816
Pro Arg Tyr Gly Gly Val Tyr Val Gly Thr Leu Ser Lys Pro Glu Val
            260                 265                 270

aag gaa gcc gtt gaa tct gct gac ttg att ttg tct gtc ggt gct ttg      864
Lys Glu Ala Val Glu Ser Ala Asp Leu Ile Leu Ser Val Gly Ala Leu
        275                 280                 285

ttg tct gat ttc aac acc ggt tct ttc tct tac tct tac aag acc aag      912
Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Lys Thr Lys
    290                 295                 300

aac att gtc gaa ttc cac tcc gac cac atg aag atc aga aac gcc act      960
Asn Ile Val Glu Phe His Ser Asp His Met Lys Ile Arg Asn Ala Thr
305                 310                 315                 320

ttc cca ggt gtc caa atg aaa ttc gtt ttg caa aag ttg ttg acc act     1008
Phe Pro Gly Val Gln Met Lys Phe Val Leu Gln Lys Leu Leu Thr Thr
                325                 330                 335

att gct gac gcc gct aag ggt tac aag cca gtt gct gtc cca gct aga     1056
Ile Ala Asp Ala Ala Lys Gly Tyr Lys Pro Val Ala Val Pro Ala Arg
            340                 345                 350

act cca gct aac gct gct gtc cca gct tct acc cca ttg aag caa gaa     1104
```

```
Thr Pro Ala Asn Ala Ala Val Pro Ala Ser Thr Pro Leu Lys Gln Glu
        355                 360                 365

tgg atg tgg aac caa ttg ggt aac ttc ttg caa gaa ggt gat gtt gtc     1152
Trp Met Trp Asn Gln Leu Gly Asn Phe Leu Gln Glu Gly Asp Val Val
    370                 375                 380

att gct gaa acc ggt acc tcc gct ttc ggt atc aac caa acc act ttc     1200
Ile Ala Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr Thr Phe
385                 390                 395                 400

cca aac aac acc tac ggt atc tct caa gtc tta tgg ggt tcc att ggt     1248
Pro Asn Asn Thr Tyr Gly Ile Ser Gln Val Leu Trp Gly Ser Ile Gly
                405                 410                 415

ttc acc act ggt gct acc ttg ggt gct gct ttc gct gct gaa gaa att     1296
Phe Thr Thr Gly Ala Thr Leu Gly Ala Ala Phe Ala Ala Glu Glu Ile
            420                 425                 430

gat cca aag aag aga gtt atc tta ttc att ggt gac ggt tct ttg caa     1344
Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln
        435                 440                 445

ttg act gtt caa gaa atc tcc acc atg atc aga tgg ggc ttg aag cca     1392
Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro
    450                 455                 460

tac ttg ttc gtc ttg aac aac gat ggt tac acc att gaa aag ttg att     1440
Tyr Leu Phe Val Leu Asn Asn Asp Gly Tyr Thr Ile Glu Lys Leu Ile
465                 470                 475                 480

cac ggt cca aag gct caa tac aac gaa att caa ggt tgg gac cac cta     1488
His Gly Pro Lys Ala Gln Tyr Asn Glu Ile Gln Gly Trp Asp His Leu
                485                 490                 495

tcc ttg ttg cca act ttc ggt gct aag gac tat gaa acc cac aga gtc     1536
Ser Leu Leu Pro Thr Phe Gly Ala Lys Asp Tyr Glu Thr His Arg Val
            500                 505                 510

gct acc acc ggt gaa tgg gac aag ttg acc caa gac aag tct ttc aac     1584
Ala Thr Thr Gly Glu Trp Asp Lys Leu Thr Gln Asp Lys Ser Phe Asn
        515                 520                 525

gac aac tct aag atc aga atg att gaa atc atg ttg cca gtc ttc gat     1632
Asp Asn Ser Lys Ile Arg Met Ile Glu Ile Met Leu Pro Val Phe Asp
    530                 535                 540

gct cca caa aac ttg gtt gaa caa gct aag ttg act gct gct acc aac     1680
Ala Pro Gln Asn Leu Val Glu Gln Ala Lys Leu Thr Ala Ala Thr Asn
545                 550                 555                 560

gct aag caa taa                                                    1692
Ala Lys Gln
```

<210> 21

<211> 563

<212> PRT

<213> Saccharomyces cerevisiae

<400> 21

```
Met Ser Glu Ile Thr Leu Gly Lys Tyr Leu Phe Glu Arg Leu Lys Gln
1               5                  10                  15

Val Asn Val Asn Thr Val Phe Gly Leu Pro Gly Asp Phe Asn Leu Ser
            20                  25                  30

Leu Leu Asp Lys Ile Tyr Glu Val Glu Gly Met Arg Trp Ala Gly Asn
        35                  40                  45

Ala Asn Glu Leu Asn Ala Ala Tyr Ala Ala Asp Gly Tyr Ala Arg Ile
    50                  55                  60

Lys Gly Met Ser Cys Ile Ile Thr Thr Phe Gly Val Gly Glu Leu Ser
65                  70                  75                  80

Ala Leu Asn Gly Ile Ala Gly Ser Tyr Ala Glu His Val Gly Val Leu
                85                  90                  95

His Val Val Gly Val Pro Ser Ile Ser Ala Gln Ala Lys Gln Leu Leu
            100                 105                 110

Leu His His Thr Leu Gly Asn Gly Asp Phe Thr Val Phe His Arg Met
        115                 120                 125

Ser Ala Asn Ile Ser Glu Thr Thr Ala Met Ile Thr Asp Ile Ala Thr
    130                 135                 140

Ala Pro Ala Glu Ile Asp Arg Cys Ile Arg Thr Thr Tyr Val Thr Gln
145                 150                 155                 160

Arg Pro Val Tyr Leu Gly Leu Pro Ala Asn Leu Val Asp Leu Asn Val
                165                 170                 175

Pro Ala Lys Leu Leu Gln Thr Pro Ile Asp Met Ser Leu Lys Pro Asn
            180                 185                 190

Asp Ala Glu Ser Glu Lys Glu Val Ile Asp Thr Ile Leu Ala Leu Val
        195                 200                 205
```

```
Lys Asp Ala Lys Asn Pro Val Ile Leu Ala Asp Ala Cys Cys Ser Arg
    210                 215                 220

His Asp Val Lys Ala Glu Thr Lys Lys Leu Ile Asp Leu Thr Gln Phe
225                 230                 235                 240

Pro Ala Phe Val Thr Pro Met Gly Lys Gly Ser Ile Asp Glu Gln His
                245                 250                 255

Pro Arg Tyr Gly Gly Val Tyr Val Gly Thr Leu Ser Lys Pro Glu Val
            260                 265                 270

Lys Glu Ala Val Glu Ser Ala Asp Leu Ile Leu Ser Val Gly Ala Leu
        275                 280                 285

Leu Ser Asp Phe Asn Thr Gly Ser Phe Ser Tyr Ser Tyr Lys Thr Lys
    290                 295                 300

Asn Ile Val Glu Phe His Ser Asp His Met Lys Ile Arg Asn Ala Thr
305                 310                 315                 320

Phe Pro Gly Val Gln Met Lys Phe Val Leu Gln Lys Leu Leu Thr Thr
                325                 330                 335

Ile Ala Asp Ala Ala Lys Gly Tyr Lys Pro Val Ala Val Pro Ala Arg
            340                 345                 350

Thr Pro Ala Asn Ala Ala Val Pro Ala Ser Thr Pro Leu Lys Gln Glu
        355                 360                 365

Trp Met Trp Asn Gln Leu Gly Asn Phe Leu Gln Glu Gly Asp Val Val
    370                 375                 380

Ile Ala Glu Thr Gly Thr Ser Ala Phe Gly Ile Asn Gln Thr Thr Phe
385                 390                 395                 400

Pro Asn Asn Thr Tyr Gly Ile Ser Gln Val Leu Trp Gly Ser Ile Gly
                405                 410                 415

Phe Thr Thr Gly Ala Thr Leu Gly Ala Ala Phe Ala Ala Glu Glu Ile
            420                 425                 430

Asp Pro Lys Lys Arg Val Ile Leu Phe Ile Gly Asp Gly Ser Leu Gln
```

435 440 445

Leu Thr Val Gln Glu Ile Ser Thr Met Ile Arg Trp Gly Leu Lys Pro
450 455 460

Tyr Leu Phe Val Leu Asn Asn Asp Gly Tyr Thr Ile Glu Lys Leu Ile
465 470 475 480

His Gly Pro Lys Ala Gln Tyr Asn Glu Ile Gln Gly Trp Asp His Leu
485 490 495

Ser Leu Leu Pro Thr Phe Gly Ala Lys Asp Tyr Glu Thr His Arg Val
500 505 510

Ala Thr Thr Gly Glu Trp Asp Lys Leu Thr Gln Asp Lys Ser Phe Asn
515 520 525

Asp Asn Ser Lys Ile Arg Met Ile Glu Ile Met Leu Pro Val Phe Asp
530 535 540

Ala Pro Gln Asn Leu Val Glu Gln Ala Lys Leu Thr Ala Ala Thr Asn
545 550 555 560

Ala Lys Gln

<210> 22

<211> 33

<212> DNA

<213> artificial sequence

<220>

<223> artificial sequence

<400> 22
tctttaatta atgggttgtc cgtcattcat ata 33

<210> 23

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> artificial sequence

<400> 23
ctaaagcttt taggccagag tggtcttgcg cg 32

<210> 24

<211> 1674

<212> DNA

<213> Acetobacter pasteurianus

<220>

<221> CDS

<222> (1)..(1674)

<223>

<400> 24
gtg acc tat act gtt ggc atg tat ctt gca gaa cgc ctt gta cag atc 48
Val Thr Tyr Thr Val Gly Met Tyr Leu Ala Glu Arg Leu Val Gln Ile
1 5 10 15

ggg ctg aag cat cac ttc gcc gtg ggc ggc gac tac aat ctc gtt ctt 96
Gly Leu Lys His His Phe Ala Val Gly Gly Asp Tyr Asn Leu Val Leu
20 25 30

ctg gat cag ttg ctc ctc aac aag gac atg aaa cag atc tat tgc tgc 144
Leu Asp Gln Leu Leu Leu Asn Lys Asp Met Lys Gln Ile Tyr Cys Cys
35 40 45

aat gag ttg aac tgt ggc ttc agc gcg gaa ggc tac gcc cgt tct aac 192
Asn Glu Leu Asn Cys Gly Phe Ser Ala Glu Gly Tyr Ala Arg Ser Asn
50 55 60

ggg gct gcg gca gcg gtt gtc acc ttc agc gtt ggc gcc att tcc gcc 240
Gly Ala Ala Ala Ala Val Val Thr Phe Ser Val Gly Ala Ile Ser Ala
65 70 75 80

atg aac gcc ctc ggc ggc gcc tat gcc gaa aac ctg ccg gtt atc ctg 288
Met Asn Ala Leu Gly Gly Ala Tyr Ala Glu Asn Leu Pro Val Ile Leu
85 90 95

att tcc ggc gcg ccc aac agc aat gat cag ggc aca ggt cat atc ctg 336
Ile Ser Gly Ala Pro Asn Ser Asn Asp Gln Gly Thr Gly His Ile Leu

```
                100                 105                 110

cat cac aca atc ggc aag acg gat tac agc tac cag ctt gaa atg gcc      384
His His Thr Ile Gly Lys Thr Asp Tyr Ser Tyr Gln Leu Glu Met Ala
        115                 120                 125

cgt cag gtc acc tgt gcc gcc gaa agc att acc gac gct cac tcc gcc      432
Arg Gln Val Thr Cys Ala Ala Glu Ser Ile Thr Asp Ala His Ser Ala
    130                 135                 140

ccg gcc aag att gac cac gtc att cgc acg gcg ctg cgc gag cgt aag      480
Pro Ala Lys Ile Asp His Val Ile Arg Thr Ala Leu Arg Glu Arg Lys
145                 150                 155                 160

ccg gcc tat ctg gac atc gcg tgc aac att gcc tcc gag ccc tgc gtg      528
Pro Ala Tyr Leu Asp Ile Ala Cys Asn Ile Ala Ser Glu Pro Cys Val
                165                 170                 175

cgg cct ggc cct gtc agc agc ctg ctg tcc gag cct gaa atc gac cac      576
Arg Pro Gly Pro Val Ser Ser Leu Leu Ser Glu Pro Glu Ile Asp His
            180                 185                 190

acg agc ctg aag gcc gca gtg gac gcc acg gtt gcc ttg ctg aaa aat      624
Thr Ser Leu Lys Ala Ala Val Asp Ala Thr Val Ala Leu Leu Lys Asn
        195                 200                 205

cgg cca gcc ccc gtc atg ctg ctg ggc agc aag ctg cgg gcc gcc aac      672
Arg Pro Ala Pro Val Met Leu Leu Gly Ser Lys Leu Arg Ala Ala Asn
    210                 215                 220

gca ctg gcc gca acc gaa acg ctg gca gac aag ctg caa tgc gcg gtg      720
Ala Leu Ala Ala Thr Glu Thr Leu Ala Asp Lys Leu Gln Cys Ala Val
225                 230                 235                 240

acc atc atg gcg gcc gcg aaa ggc ttt ttc ccc gaa gac cac gcg ggt      768
Thr Ile Met Ala Ala Ala Lys Gly Phe Phe Pro Glu Asp His Ala Gly
                245                 250                 255

ttc cgc ggc ctg tac tgg ggc gaa gtc tcg aac ccc ggc gtg cag gaa      816
Phe Arg Gly Leu Tyr Trp Gly Glu Val Ser Asn Pro Gly Val Gln Glu
            260                 265                 270

ctg gtg gag acc tcc gac gca ctg ctg tgc atc gcc ccc gta ttc aac      864
Leu Val Glu Thr Ser Asp Ala Leu Leu Cys Ile Ala Pro Val Phe Asn
        275                 280                 285

gac tat tca aca gtc ggc tgg tcg ggc atg ccc aag ggc ccc aat gtg      912
Asp Tyr Ser Thr Val Gly Trp Ser Gly Met Pro Lys Gly Pro Asn Val
    290                 295                 300

att ctg gct gag ccc gac cgc gta acg gtc gat ggc cgc gcc tat gac      960
Ile Leu Ala Glu Pro Asp Arg Val Thr Val Asp Gly Arg Ala Tyr Asp
305                 310                 315                 320

ggc ttt acc ctg cgc gcc ttc ctg cag gct ctg gcg gaa aaa gcc ccc     1008
Gly Phe Thr Leu Arg Ala Phe Leu Gln Ala Leu Ala Glu Lys Ala Pro
                325                 330                 335

gcg cgc ccg gcc tcc gca cag aaa agc agc gtc ccg acg tgc tcg ctc     1056
```

```
Ala Arg Pro Ala Ser Ala Gln Lys Ser Ser Val Pro Thr Cys Ser Leu
            340                 345                 350

acc gcg aca tcc gat gaa gcc ggt ctg acg aat gac gaa atc gtc cgt    1104
Thr Ala Thr Ser Asp Glu Ala Gly Leu Thr Asn Asp Glu Ile Val Arg
        355                 360                 365

cat atc aac gcc ctg ctg aca tca aac acg acg ctg gtg gca gaa acc    1152
His Ile Asn Ala Leu Leu Thr Ser Asn Thr Thr Leu Val Ala Glu Thr
    370                 375                 380

ggc gat tca tgg ttc aat gcc atg cgc atg acc ctg gcc ggt gcg cgc    1200
Gly Asp Ser Trp Phe Asn Ala Met Arg Met Thr Leu Ala Gly Ala Arg
385                 390                 395                 400

gtg gaa ctg gaa atg cag tgg ggc cat atc ggc tgg tcc gtg ccc tcc    1248
Val Glu Leu Glu Met Gln Trp Gly His Ile Gly Trp Ser Val Pro Ser
                405                 410                 415

gcg ttc ggc aat gcc atg ggc tcg cag gac cgc cag cat gtg gtg atg    1296
Ala Phe Gly Asn Ala Met Gly Ser Gln Asp Arg Gln His Val Val Met
            420                 425                 430

gta ggc gat ggc tcc ttc cag ctt acc gcg cag gaa gtg gct cag atg    1344
Val Gly Asp Gly Ser Phe Gln Leu Thr Ala Gln Glu Val Ala Gln Met
        435                 440                 445

gtg cgc tac gaa ctg ccc gtc att atc ttt ctg atc aac aac cgt ggc    1392
Val Arg Tyr Glu Leu Pro Val Ile Ile Phe Leu Ile Asn Asn Arg Gly
    450                 455                 460

tat gtc att gaa atc gcc att cat gac ggc ccg tac aac tat atc aag    1440
Tyr Val Ile Glu Ile Ala Ile His Asp Gly Pro Tyr Asn Tyr Ile Lys
465                 470                 475                 480

aac tgg gat tac gcc ggc ctg atg gaa gtc ttc aac gcc gga gaa ggc    1488
Asn Trp Asp Tyr Ala Gly Leu Met Glu Val Phe Asn Ala Gly Glu Gly
                485                 490                 495

cat gga ctt ggc ctg aaa gcc acc acc ccg aag gaa ctg aca gaa gcc    1536
His Gly Leu Gly Leu Lys Ala Thr Thr Pro Lys Glu Leu Thr Glu Ala
            500                 505                 510

atc gcc agg gca aaa gcc aat acc cgc ggc ccg acg ctg atc gaa tgc    1584
Ile Ala Arg Ala Lys Ala Asn Thr Arg Gly Pro Thr Leu Ile Glu Cys
        515                 520                 525

cag atc gac cgc acg gac tgc acg gat atg ctg gtt caa tgg ggc cgc    1632
Gln Ile Asp Arg Thr Asp Cys Thr Asp Met Leu Val Gln Trp Gly Arg
    530                 535                 540

aag gtt gcc tca acc aac gcg cgc aag acc act ctg gcc tga             1674
Lys Val Ala Ser Thr Asn Ala Arg Lys Thr Thr Leu Ala
545                 550                 555
```

<210> 25

<211> 557

<212> PRT

<213> Acetobacter pasteurianus

<400> 25

```
Val Thr Tyr Thr Val Gly Met Tyr Leu Ala Glu Arg Leu Val Gln Ile
1               5                  10                  15

Gly Leu Lys His His Phe Ala Val Gly Gly Asp Tyr Asn Leu Val Leu
            20                  25                  30

Leu Asp Gln Leu Leu Leu Asn Lys Asp Met Lys Gln Ile Tyr Cys Cys
        35                  40                  45

Asn Glu Leu Asn Cys Gly Phe Ser Ala Glu Gly Tyr Ala Arg Ser Asn
    50                  55                  60

Gly Ala Ala Ala Ala Val Val Thr Phe Ser Val Gly Ala Ile Ser Ala
65                  70                  75                  80

Met Asn Ala Leu Gly Gly Ala Tyr Ala Glu Asn Leu Pro Val Ile Leu
                85                  90                  95

Ile Ser Gly Ala Pro Asn Ser Asn Asp Gln Gly Thr Gly His Ile Leu
            100                 105                 110

His His Thr Ile Gly Lys Thr Asp Tyr Ser Tyr Gln Leu Glu Met Ala
        115                 120                 125

Arg Gln Val Thr Cys Ala Ala Glu Ser Ile Thr Asp Ala His Ser Ala
    130                 135                 140

Pro Ala Lys Ile Asp His Val Ile Arg Thr Ala Leu Arg Glu Arg Lys
145                 150                 155                 160

Pro Ala Tyr Leu Asp Ile Ala Cys Asn Ile Ala Ser Glu Pro Cys Val
                165                 170                 175

Arg Pro Gly Pro Val Ser Ser Leu Leu Ser Glu Pro Glu Ile Asp His
            180                 185                 190

Thr Ser Leu Lys Ala Ala Val Asp Ala Thr Val Ala Leu Leu Lys Asn
        195                 200                 205
```

```
Arg Pro Ala Pro Val Met Leu Leu Gly Ser Lys Leu Arg Ala Ala Asn
    210                 215                 220

Ala Leu Ala Ala Thr Glu Thr Leu Ala Asp Lys Leu Gln Cys Ala Val
225                 230                 235                 240

Thr Ile Met Ala Ala Ala Lys Gly Phe Phe Pro Glu Asp His Ala Gly
                245                 250                 255

Phe Arg Gly Leu Tyr Trp Gly Glu Val Ser Asn Pro Gly Val Gln Glu
            260                 265                 270

Leu Val Glu Thr Ser Asp Ala Leu Leu Cys Ile Ala Pro Val Phe Asn
        275                 280                 285

Asp Tyr Ser Thr Val Gly Trp Ser Gly Met Pro Lys Gly Pro Asn Val
    290                 295                 300

Ile Leu Ala Glu Pro Asp Arg Val Thr Val Asp Gly Arg Ala Tyr Asp
305                 310                 315                 320

Gly Phe Thr Leu Arg Ala Phe Leu Gln Ala Leu Ala Glu Lys Ala Pro
                325                 330                 335

Ala Arg Pro Ala Ser Ala Gln Lys Ser Ser Val Pro Thr Cys Ser Leu
            340                 345                 350

Thr Ala Thr Ser Asp Glu Ala Gly Leu Thr Asn Asp Glu Ile Val Arg
        355                 360                 365

His Ile Asn Ala Leu Leu Thr Ser Asn Thr Thr Leu Val Ala Glu Thr
    370                 375                 380

Gly Asp Ser Trp Phe Asn Ala Met Arg Met Thr Leu Ala Gly Ala Arg
385                 390                 395                 400

Val Glu Leu Glu Met Gln Trp Gly His Ile Gly Trp Ser Val Pro Ser
                405                 410                 415

Ala Phe Gly Asn Ala Met Gly Ser Gln Asp Arg Gln His Val Val Met
            420                 425                 430

Val Gly Asp Gly Ser Phe Gln Leu Thr Ala Gln Glu Val Ala Gln Met
```

435 440 445

Val Arg Tyr Glu Leu Pro Val Ile Ile Phe Leu Ile Asn Asn Arg Gly
450 455 460

Tyr Val Ile Glu Ile Ala Ile His Asp Gly Pro Tyr Asn Tyr Ile Lys
465 470 475 480

Asn Trp Asp Tyr Ala Gly Leu Met Glu Val Phe Asn Ala Gly Glu Gly
485 490 495

His Gly Leu Gly Leu Lys Ala Thr Thr Pro Lys Glu Leu Thr Glu Ala
500 505 510

Ile Ala Arg Ala Lys Ala Asn Thr Arg Gly Pro Thr Leu Ile Glu Cys
515 520 525

Gln Ile Asp Arg Thr Asp Cys Thr Asp Met Leu Val Gln Trp Gly Arg
530 535 540

Lys Val Ala Ser Thr Asn Ala Arg Lys Thr Thr Leu Ala
545 550 555

<210> 26

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> artificial sequence

<400> 26
atcttaatta atgtataccg ttggtatgta ct 32

<210> 27

<211> 34

<212> DNA

<213> artificial sequence

<220>

<223> artificial sequence

<400> 27
tatgcggccg cttacgcttg tggtttgcga gagt 34

<210> 28

<211> 1671

<212> DNA

<213> Zymobacter palmae

<220>

<221> CDS

<222> (1)..(1671)

<223>

<400> 28
atg tat acc gtt ggt atg tac ttg gca gaa cgc cta gcc cag atc ggc 48
Met Tyr Thr Val Gly Met Tyr Leu Ala Glu Arg Leu Ala Gln Ile Gly
1 5 10 15

ctg aaa cac cac ttt gcc gtg gcc ggt gac tac aac ctg gtg ttg ctt 96
Leu Lys His His Phe Ala Val Ala Gly Asp Tyr Asn Leu Val Leu Leu
20 25 30

gat cag ctc ctg ctg aac aaa gac atg gag cag gtc tac tgc tgt aac 144
Asp Gln Leu Leu Leu Asn Lys Asp Met Glu Gln Val Tyr Cys Cys Asn
35 40 45

gaa ctt aac tgc ggc ttt agc gcc gaa ggt tac gct cgt gca cgt ggt 192
Glu Leu Asn Cys Gly Phe Ser Ala Glu Gly Tyr Ala Arg Ala Arg Gly
50 55 60

gcc gcc gct gcc atc gtc acg ttc agc gta ggt gct atc tct gca atg 240
Ala Ala Ala Ala Ile Val Thr Phe Ser Val Gly Ala Ile Ser Ala Met
65 70 75 80

aac gcc atc ggt ggc gcc tat gca gaa aac ctg ccg gtc atc ctg atc 288
Asn Ala Ile Gly Gly Ala Tyr Ala Glu Asn Leu Pro Val Ile Leu Ile
85 90 95

tct ggc tca ccg aac acc aat gac tac ggc aca ggc cac atc ctg cac 336
Ser Gly Ser Pro Asn Thr Asn Asp Tyr Gly Thr Gly His Ile Leu His
100 105 110

cac acc att ggt act act gac tat aac tat cag ctg gaa atg gta aaa 384
His Thr Ile Gly Thr Thr Asp Tyr Asn Tyr Gln Leu Glu Met Val Lys

```
        115                 120                 125

cac gtt acc tgc gca cgt gaa agc atc gtt tct gcc gaa gaa gca ccg      432
His Val Thr Cys Ala Arg Glu Ser Ile Val Ser Ala Glu Glu Ala Pro
    130                 135                 140

gca aaa atc gac cac gtc atc cgt acg gct cta cgt gaa cgc aaa ccg      480
Ala Lys Ile Asp His Val Ile Arg Thr Ala Leu Arg Glu Arg Lys Pro
145                 150                 155                 160

gct tat ctg gaa atc gca tgc aac gtc gct ggc gct gaa tgt gtt cgt      528
Ala Tyr Leu Glu Ile Ala Cys Asn Val Ala Gly Ala Glu Cys Val Arg
                165                 170                 175

ccg ggc ccg atc aat agc ctg ctg cgt gaa ctc gaa gtt gac cag acc      576
Pro Gly Pro Ile Asn Ser Leu Leu Arg Glu Leu Glu Val Asp Gln Thr
            180                 185                 190

agt gtc act gcc gct gta gat gcc gcc gta gaa tgg ctg cag gac cgc      624
Ser Val Thr Ala Ala Val Asp Ala Ala Val Glu Trp Leu Gln Asp Arg
        195                 200                 205

cag aac gtc gtc atg ctg gtc ggt agc aaa ctg cgt gcc gct gcc gct      672
Gln Asn Val Val Met Leu Val Gly Ser Lys Leu Arg Ala Ala Ala Ala
    210                 215                 220

gaa aaa cag gct gtt gcc cta gcg gac cgc ctg ggc tgc gct gtc acg      720
Glu Lys Gln Ala Val Ala Leu Ala Asp Arg Leu Gly Cys Ala Val Thr
225                 230                 235                 240

atc atg gct gcc gaa aaa ggc ttc ttc ccg gaa gat cat ccg aac ttc      768
Ile Met Ala Ala Glu Lys Gly Phe Phe Pro Glu Asp His Pro Asn Phe
                245                 250                 255

cgc ggc ctg tac tgg ggt gaa gtc agc tcc gaa ggt gca cag gaa ctg      816
Arg Gly Leu Tyr Trp Gly Glu Val Ser Ser Glu Gly Ala Gln Glu Leu
            260                 265                 270

gtt gaa aac gcc gat gcc atc ctg tgt ctg gca ccg gta ttc aac gac      864
Val Glu Asn Ala Asp Ala Ile Leu Cys Leu Ala Pro Val Phe Asn Asp
        275                 280                 285

tat gct acc gtt ggc tgg aac tcc tgg ccg aaa ggc gac aat gtc atg      912
Tyr Ala Thr Val Gly Trp Asn Ser Trp Pro Lys Gly Asp Asn Val Met
    290                 295                 300

gtc atg gac acc gac cgc gtc act ttc gca gga cag tcc ttc gaa ggt      960
Val Met Asp Thr Asp Arg Val Thr Phe Ala Gly Gln Ser Phe Glu Gly
305                 310                 315                 320

ctg tca ttg agc acc ttc gcc gca gca ctg gct gag aaa gca cct tct     1008
Leu Ser Leu Ser Thr Phe Ala Ala Ala Leu Ala Glu Lys Ala Pro Ser
                325                 330                 335

cgc ccg gca acg act caa ggc act caa gca ccg gta ctg ggt att gag     1056
Arg Pro Ala Thr Thr Gln Gly Thr Gln Ala Pro Val Leu Gly Ile Glu
            340                 345                 350

gcc gca gag ccc aat gca ccg ctg acc aat gac gaa atg acg cgt cag     1104
```

```
Ala Ala Glu Pro Asn Ala Pro Leu Thr Asn Asp Glu Met Thr Arg Gln
        355                 360                 365

atc cag tcg ctg atc act tcc gac act act ctg aca gca gaa aca ggt     1152
Ile Gln Ser Leu Ile Thr Ser Asp Thr Thr Leu Thr Ala Glu Thr Gly
    370                 375                 380

gac tct tgg ttc aac gct tct cgc atg ccg att cct ggc ggt gct cgt     1200
Asp Ser Trp Phe Asn Ala Ser Arg Met Pro Ile Pro Gly Gly Ala Arg
385                 390                 395                 400

gtc gaa ctg gaa atg caa tgg ggt cat atc ggt tgg tcc gta cct tct     1248
Val Glu Leu Glu Met Gln Trp Gly His Ile Gly Trp Ser Val Pro Ser
                405                 410                 415

gca ttc ggt aac gcc gtt ggt tct ccg gag cgt cgc cac atc atg atg     1296
Ala Phe Gly Asn Ala Val Gly Ser Pro Glu Arg Arg His Ile Met Met
            420                 425                 430

gtc ggt gat ggc tct ttc cag ctg act gct caa gaa gtt gct cag atg     1344
Val Gly Asp Gly Ser Phe Gln Leu Thr Ala Gln Glu Val Ala Gln Met
        435                 440                 445

atc cgc tat gaa atc ccg gtc atc atc ttc ctg atc aac aac cgc ggt     1392
Ile Arg Tyr Glu Ile Pro Val Ile Ile Phe Leu Ile Asn Asn Arg Gly
    450                 455                 460

tac gtc atc gaa atc gct atc cat gac ggc cct tac aac tac atc aaa     1440
Tyr Val Ile Glu Ile Ala Ile His Asp Gly Pro Tyr Asn Tyr Ile Lys
465                 470                 475                 480

aac tgg aac tac gct ggc ctg atc gac gtc ttc aat gac gaa gat ggt     1488
Asn Trp Asn Tyr Ala Gly Leu Ile Asp Val Phe Asn Asp Glu Asp Gly
                485                 490                 495

cat ggc ctg ggt ctg aaa gct tct act ggt gca gaa cta gaa ggc gct     1536
His Gly Leu Gly Leu Lys Ala Ser Thr Gly Ala Glu Leu Glu Gly Ala
            500                 505                 510

atc aag aaa gca ctc gac aat cgt cgc ggt ccg acg ctg atc gaa tgt     1584
Ile Lys Lys Ala Leu Asp Asn Arg Arg Gly Pro Thr Leu Ile Glu Cys
        515                 520                 525

aac atc gct cag gac gac tgc act gaa acc ctg att gct tgg ggt aaa     1632
Asn Ile Ala Gln Asp Asp Cys Thr Glu Thr Leu Ile Ala Trp Gly Lys
    530                 535                 540

cgt gta gca gct acc aac tct cgc aaa cca caa gcg taa                 1671
Arg Val Ala Ala Thr Asn Ser Arg Lys Pro Gln Ala
545                 550                 555
```

<210> 29

<211> 556

<212> PRT

<213> Zymobacter palmae

<400> 29

```
Met Tyr Thr Val Gly Met Tyr Leu Ala Glu Arg Leu Ala Gln Ile Gly
1               5                  10                  15

Leu Lys His His Phe Ala Val Ala Gly Asp Tyr Asn Leu Val Leu Leu
            20                  25                  30

Asp Gln Leu Leu Leu Asn Lys Asp Met Glu Gln Val Tyr Cys Cys Asn
        35                  40                  45

Glu Leu Asn Cys Gly Phe Ser Ala Glu Gly Tyr Ala Arg Ala Arg Gly
    50                  55                  60

Ala Ala Ala Ala Ile Val Thr Phe Ser Val Gly Ala Ile Ser Ala Met
65                  70                  75                  80

Asn Ala Ile Gly Gly Ala Tyr Ala Glu Asn Leu Pro Val Ile Leu Ile
                85                  90                  95

Ser Gly Ser Pro Asn Thr Asn Asp Tyr Gly Thr Gly His Ile Leu His
            100                 105                 110

His Thr Ile Gly Thr Thr Asp Tyr Asn Tyr Gln Leu Glu Met Val Lys
        115                 120                 125

His Val Thr Cys Ala Arg Glu Ser Ile Val Ser Ala Glu Glu Ala Pro
    130                 135                 140

Ala Lys Ile Asp His Val Ile Arg Thr Ala Leu Arg Glu Arg Lys Pro
145                 150                 155                 160

Ala Tyr Leu Glu Ile Ala Cys Asn Val Ala Gly Ala Glu Cys Val Arg
                165                 170                 175

Pro Gly Pro Ile Asn Ser Leu Leu Arg Glu Leu Glu Val Asp Gln Thr
            180                 185                 190

Ser Val Thr Ala Ala Val Asp Ala Ala Val Glu Trp Leu Gln Asp Arg
        195                 200                 205

Gln Asn Val Val Met Leu Val Gly Ser Lys Leu Arg Ala Ala Ala Ala
    210                 215                 220
```

```
Glu Lys Gln Ala Val Ala Leu Ala Asp Arg Leu Gly Cys Ala Val Thr
225                 230                 235                 240

Ile Met Ala Ala Glu Lys Gly Phe Phe Pro Glu Asp His Pro Asn Phe
                245                 250                 255

Arg Gly Leu Tyr Trp Gly Glu Val Ser Ser Glu Gly Ala Gln Glu Leu
            260                 265                 270

Val Glu Asn Ala Asp Ala Ile Leu Cys Leu Ala Pro Val Phe Asn Asp
        275                 280                 285

Tyr Ala Thr Val Gly Trp Asn Ser Trp Pro Lys Gly Asp Asn Val Met
    290                 295                 300

Val Met Asp Thr Asp Arg Val Thr Phe Ala Gly Gln Ser Phe Glu Gly
305                 310                 315                 320

Leu Ser Leu Ser Thr Phe Ala Ala Ala Leu Ala Glu Lys Ala Pro Ser
                325                 330                 335

Arg Pro Ala Thr Thr Gln Gly Thr Gln Ala Pro Val Leu Gly Ile Glu
            340                 345                 350

Ala Ala Glu Pro Asn Ala Pro Leu Thr Asn Asp Glu Met Thr Arg Gln
        355                 360                 365

Ile Gln Ser Leu Ile Thr Ser Asp Thr Thr Leu Thr Ala Glu Thr Gly
    370                 375                 380

Asp Ser Trp Phe Asn Ala Ser Arg Met Pro Ile Pro Gly Gly Ala Arg
385                 390                 395                 400

Val Glu Leu Glu Met Gln Trp Gly His Ile Gly Trp Ser Val Pro Ser
                405                 410                 415

Ala Phe Gly Asn Ala Val Gly Ser Pro Glu Arg Arg His Ile Met Met
            420                 425                 430

Val Gly Asp Gly Ser Phe Gln Leu Thr Ala Gln Glu Val Ala Gln Met
        435                 440                 445

Ile Arg Tyr Glu Ile Pro Val Ile Ile Phe Leu Ile Asn Asn Arg Gly
```

450 455 460

Tyr Val Ile Glu Ile Ala Ile His Asp Gly Pro Tyr Asn Tyr Ile Lys
465 470 475 480

Asn Trp Asn Tyr Ala Gly Leu Ile Asp Val Phe Asn Asp Glu Asp Gly
485 490 495

His Gly Leu Gly Leu Lys Ala Ser Thr Gly Ala Glu Leu Glu Gly Ala
500 505 510

Ile Lys Lys Ala Leu Asp Asn Arg Arg Gly Pro Thr Leu Ile Glu Cys
515 520 525

Asn Ile Ala Gln Asp Asp Cys Thr Glu Thr Leu Ile Ala Trp Gly Lys
530 535 540

Arg Val Ala Ala Thr Asn Ser Arg Lys Pro Gln Ala
545 550 555

<210> 30

<211> 32

<212> DNA

<213> artificial sequence

<400> 30
ctattaatta atggcttcgg tacacggcac ca 32

<210> 31

<211> 34

<212> DNA

<213> artificial sequence

<400> 31
tatgcggccg cttacttcac cgggcttacg gtgc 34

<210> 32

<211> 1587

<212> DNA

<213> Pseudomonas putida

<220>

<221> CDS

<222> (1)..(1584)

<223>

<400> 32

```
atg gct tcg gta cac ggc acc aca tac gaa ctc ttg cga cgt caa ggc       48
Met Ala Ser Val His Gly Thr Thr Tyr Glu Leu Leu Arg Arg Gln Gly
1               5                   10                  15

atc gat acg gtc ttc ggc aat cct ggc tcg aac gag ctc ccg ttt ttg       96
Ile Asp Thr Val Phe Gly Asn Pro Gly Ser Asn Glu Leu Pro Phe Leu
            20                  25                  30

aag gac ttt cca gag gac ttt cga tac atc ctg gct ttg cag gaa gcg      144
Lys Asp Phe Pro Glu Asp Phe Arg Tyr Ile Leu Ala Leu Gln Glu Ala
        35                  40                  45

tgt gtg gtg ggc att gca gac ggc tat gcg caa gcc agt cgg aag ccg      192
Cys Val Val Gly Ile Ala Asp Gly Tyr Ala Gln Ala Ser Arg Lys Pro
    50                  55                  60

gct ttc att aac ctg cat tct gct gct ggt acc ggc aat gct atg ggt      240
Ala Phe Ile Asn Leu His Ser Ala Ala Gly Thr Gly Asn Ala Met Gly
65                  70                  75                  80

gca ctc agt aac gcc tgg aac tca cat tcc ccg ctg atc gtc act gcc      288
Ala Leu Ser Asn Ala Trp Asn Ser His Ser Pro Leu Ile Val Thr Ala
                85                  90                  95

ggc cag cag acc agg gcg atg att ggc gtt gaa gct ctg ctg acc aac      336
Gly Gln Gln Thr Arg Ala Met Ile Gly Val Glu Ala Leu Leu Thr Asn
            100                 105                 110

gtc gat gcc gcc aac ctg cca cga cca ctt gtc aaa tgg agc tac gag      384
Val Asp Ala Ala Asn Leu Pro Arg Pro Leu Val Lys Trp Ser Tyr Glu
        115                 120                 125

ccc gca agc gca gca gaa gtc cct cat gcg atg agc agg gct atc cat      432
Pro Ala Ser Ala Ala Glu Val Pro His Ala Met Ser Arg Ala Ile His
    130                 135                 140

atg gca agc atg gcg cca caa ggc cct gtc tat ctt tcg gtg cca tat      480
Met Ala Ser Met Ala Pro Gln Gly Pro Val Tyr Leu Ser Val Pro Tyr
145                 150                 155                 160

gac gat tgg gat aag gat gct gat cct cag tcc cac cac ctt ttt gat      528
Asp Asp Trp Asp Lys Asp Ala Asp Pro Gln Ser His His Leu Phe Asp
```

```
                165                 170                 175

cgc cat gtc agt tca tca gta cgc ctg aac gac cag gat ctc gat att      576
Arg His Val Ser Ser Ser Val Arg Leu Asn Asp Gln Asp Leu Asp Ile
            180                 185                 190

ctg gtg aaa gct ctc aac agc gca tcc aac ccg gcg atc gtc ctg ggc      624
Leu Val Lys Ala Leu Asn Ser Ala Ser Asn Pro Ala Ile Val Leu Gly
        195                 200                 205

ccg gac gtc gac gca gca aat gcg aac gca gac tgc gtc atg ttg gcc      672
Pro Asp Val Asp Ala Ala Asn Ala Asn Ala Asp Cys Val Met Leu Ala
    210                 215                 220

gaa cgc ctc aaa gct ccg gtt tgg gtt gcg cca tcc gct cca cgc tgc      720
Glu Arg Leu Lys Ala Pro Val Trp Val Ala Pro Ser Ala Pro Arg Cys
225                 230                 235                 240

cca ttc cct acc cgt cat cct tgc ttc cgt gga ttg atg cca gct ggc      768
Pro Phe Pro Thr Arg His Pro Cys Phe Arg Gly Leu Met Pro Ala Gly
                245                 250                 255

atc gca gcg att tct cag ctg ctc gaa ggt cac gat gtg gtt ttg gta      816
Ile Ala Ala Ile Ser Gln Leu Leu Glu Gly His Asp Val Val Leu Val
            260                 265                 270

atc ggc gct cca gtg ttc cgt tac cac caa tac gac cca ggt caa tat      864
Ile Gly Ala Pro Val Phe Arg Tyr His Gln Tyr Asp Pro Gly Gln Tyr
        275                 280                 285

ctc aaa cct ggc acg cga ttg att tcg gtg acc tgc gac ccg ctc gaa      912
Leu Lys Pro Gly Thr Arg Leu Ile Ser Val Thr Cys Asp Pro Leu Glu
    290                 295                 300

gct gca cgc gcg cca atg ggc gat gcg atc gtg gca gac att ggt gcg      960
Ala Ala Arg Ala Pro Met Gly Asp Ala Ile Val Ala Asp Ile Gly Ala
305                 310                 315                 320

atg gct agc gct ctt gcc aac ttg gtt gaa gag agc agc cgc cag ctc     1008
Met Ala Ser Ala Leu Ala Asn Leu Val Glu Glu Ser Ser Arg Gln Leu
                325                 330                 335

cca act gca gct ccg gaa ccc gcg aag gtt gac caa gac gct ggc cga     1056
Pro Thr Ala Ala Pro Glu Pro Ala Lys Val Asp Gln Asp Ala Gly Arg
            340                 345                 350

ctt cac cca gag aca gtg ttc gac aca ctg aac gac atg gcc ccg gag     1104
Leu His Pro Glu Thr Val Phe Asp Thr Leu Asn Asp Met Ala Pro Glu
        355                 360                 365

aat gcg att tac ctg aac gag tcg act tca acg acc gcc caa atg tgg     1152
Asn Ala Ile Tyr Leu Asn Glu Ser Thr Ser Thr Thr Ala Gln Met Trp
    370                 375                 380

cag cgc ctg aac atg cgc aac cct ggt agc tac tac ttc tgt gca gct     1200
Gln Arg Leu Asn Met Arg Asn Pro Gly Ser Tyr Tyr Phe Cys Ala Ala
385                 390                 395                 400

ggc gga ctg ggc ttc gcc ctg cct gca gca att ggc gtt caa ctc gca     1248
```

```
Gly Gly Leu Gly Phe Ala Leu Pro Ala Ala Ile Gly Val Gln Leu Ala
                405                 410                 415

gaa ccc gag cga caa gtc atc gcc gtc att ggc gac gga tcg gcg aac     1296
Glu Pro Glu Arg Gln Val Ile Ala Val Ile Gly Asp Gly Ser Ala Asn
            420                 425                 430

tac agc att agt gcg ttg tgg act gca gct cag tac aac atc ccc act     1344
Tyr Ser Ile Ser Ala Leu Trp Thr Ala Ala Gln Tyr Asn Ile Pro Thr
        435                 440                 445

atc ttc gtg atc atg aac aac ggc acc tac ggt gcg ttg cga tgg ttt     1392
Ile Phe Val Ile Met Asn Asn Gly Thr Tyr Gly Ala Leu Arg Trp Phe
    450                 455                 460

gcc ggc gtt ctc gaa gca gaa aac gtt cct ggg ctg gat gtg cca ggg     1440
Ala Gly Val Leu Glu Ala Glu Asn Val Pro Gly Leu Asp Val Pro Gly
465                 470                 475                 480

atc gac ttc cgc gca ctc gcc aag ggc tat ggt gtc caa gcg ctg aaa     1488
Ile Asp Phe Arg Ala Leu Ala Lys Gly Tyr Gly Val Gln Ala Leu Lys
                485                 490                 495

gcc gac aac ctt gag cag ctc aag ggt tcg cta caa gaa gcg ctt tct     1536
Ala Asp Asn Leu Glu Gln Leu Lys Gly Ser Leu Gln Glu Ala Leu Ser
            500                 505                 510

gcc aaa ggc ccg gta ctt atc gaa gta agc acc gta agc ccg gtg aag     1584
Ala Lys Gly Pro Val Leu Ile Glu Val Ser Thr Val Ser Pro Val Lys
        515                 520                 525

tga                                                                1587


<210>  33

<211>  528

<212>  PRT

<213>  Pseudomonas putida


<400>  33

Met Ala Ser Val His Gly Thr Thr Tyr Glu Leu Leu Arg Arg Gln Gly
1               5                   10                  15

Ile Asp Thr Val Phe Gly Asn Pro Gly Ser Asn Glu Leu Pro Phe Leu
            20                  25                  30

Lys Asp Phe Pro Glu Asp Phe Arg Tyr Ile Leu Ala Leu Gln Glu Ala
        35                  40                  45

Cys Val Val Gly Ile Ala Asp Gly Tyr Ala Gln Ala Ser Arg Lys Pro
```

```
    50                  55                  60

Ala Phe Ile Asn Leu His Ser Ala Ala Gly Thr Gly Asn Ala Met Gly
65                  70                  75                  80

Ala Leu Ser Asn Ala Trp Asn Ser His Ser Pro Leu Ile Val Thr Ala
                85                  90                  95

Gly Gln Gln Thr Arg Ala Met Ile Gly Val Glu Ala Leu Leu Thr Asn
            100                 105                 110

Val Asp Ala Ala Asn Leu Pro Arg Pro Leu Val Lys Trp Ser Tyr Glu
        115                 120                 125

Pro Ala Ser Ala Ala Glu Val Pro His Ala Met Ser Arg Ala Ile His
    130                 135                 140

Met Ala Ser Met Ala Pro Gln Gly Pro Val Tyr Leu Ser Val Pro Tyr
145                 150                 155                 160

Asp Asp Trp Asp Lys Asp Ala Asp Pro Gln Ser His His Leu Phe Asp
                165                 170                 175

Arg His Val Ser Ser Ser Val Arg Leu Asn Asp Gln Asp Leu Asp Ile
            180                 185                 190

Leu Val Lys Ala Leu Asn Ser Ala Ser Asn Pro Ala Ile Val Leu Gly
        195                 200                 205

Pro Asp Val Asp Ala Ala Asn Ala Asn Ala Asp Cys Val Met Leu Ala
    210                 215                 220

Glu Arg Leu Lys Ala Pro Val Trp Val Ala Pro Ser Ala Pro Arg Cys
225                 230                 235                 240

Pro Phe Pro Thr Arg His Pro Cys Phe Arg Gly Leu Met Pro Ala Gly
                245                 250                 255

Ile Ala Ala Ile Ser Gln Leu Leu Glu Gly His Asp Val Val Leu Val
            260                 265                 270

Ile Gly Ala Pro Val Phe Arg Tyr His Gln Tyr Asp Pro Gly Gln Tyr
        275                 280                 285
```

```
Leu Lys Pro Gly Thr Arg Leu Ile Ser Val Thr Cys Asp Pro Leu Glu
    290                 295                 300

Ala Ala Arg Ala Pro Met Gly Asp Ala Ile Val Ala Asp Ile Gly Ala
305                 310                 315                 320

Met Ala Ser Ala Leu Ala Asn Leu Val Glu Glu Ser Ser Arg Gln Leu
                325                 330                 335

Pro Thr Ala Ala Pro Glu Pro Ala Lys Val Asp Gln Asp Ala Gly Arg
            340                 345                 350

Leu His Pro Glu Thr Val Phe Asp Thr Leu Asn Asp Met Ala Pro Glu
        355                 360                 365

Asn Ala Ile Tyr Leu Asn Glu Ser Thr Ser Thr Thr Ala Gln Met Trp
    370                 375                 380

Gln Arg Leu Asn Met Arg Asn Pro Gly Ser Tyr Tyr Phe Cys Ala Ala
385                 390                 395                 400

Gly Gly Leu Gly Phe Ala Leu Pro Ala Ala Ile Gly Val Gln Leu Ala
                405                 410                 415

Glu Pro Glu Arg Gln Val Ile Ala Val Ile Gly Asp Gly Ser Ala Asn
            420                 425                 430

Tyr Ser Ile Ser Ala Leu Trp Thr Ala Ala Gln Tyr Asn Ile Pro Thr
        435                 440                 445

Ile Phe Val Ile Met Asn Asn Gly Thr Tyr Gly Ala Leu Arg Trp Phe
    450                 455                 460

Ala Gly Val Leu Glu Ala Glu Asn Val Pro Gly Leu Asp Val Pro Gly
465                 470                 475                 480

Ile Asp Phe Arg Ala Leu Ala Lys Gly Tyr Gly Val Gln Ala Leu Lys
                485                 490                 495

Ala Asp Asn Leu Glu Gln Leu Lys Gly Ser Leu Gln Glu Ala Leu Ser
            500                 505                 510

Ala Lys Gly Pro Val Leu Ile Glu Val Ser Thr Val Ser Pro Val Lys
        515                 520                 525
```

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP2004/006848

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07H7/027 C07H1/00 C12P7/42 C12N9/88 C12N1/20

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07H C12P C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BEILSTEIN Data, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SHELTON M C ET AL: "2-Keto-3-deoxy-6-phosphogluconate aldolases as catalysts for stereocontrolled carbon-carbon bond formation" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, AMERICAN CHEMICAL SOCIETY, WASHINGTON, DC, US, vol. 118, no. 9, 6 March 1996 (1996-03-06), pages 2117-2125, XP002263455 ISSN: 0002-7863 Scheme 6 -----	1
A	US 5 872 247 A (DUFLOT PIERRICK ET AL) 16 February 1999 (1999-02-16) exemples ----- -/--	1

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 November 2004	29/11/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	de Nooy, A

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP2004/006848

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 846 794 A (DELOBEAU DIDIER ET AL) 8 December 1998 (1998-12-08) exemple 3	1
A	H. KILIANI, H. NAEGELL: "Ueber Meta- und Parasaccharin" CHEM. BER., vol. 35, 1902, pages 3528-3533, XP002267849 page 3531 - page 3532	13
A	WO 01/14566 A (IHLENFELDT HANS GEORG ; PHARMA WALDHOF GMBH & CO KG (DE); TISCHER WILH) 1 March 2001 (2001-03-01) the whole document	1,20,41

Form PCT/ISA/210 (continuation of second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

formation on patent family members

International Application No
PCT/EP2004/006848

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5872247 A	16-02-1999	FR 2749306 A1	05-12-1997
		AT 203996 T	15-08-2001
		CA 2206390 A1	03-12-1997
		DE 69705995 D1	13-09-2001
		DE 69705995 T2	04-04-2002
		DK 811632 T3	12-11-2001
		EP 0811632 A1	10-12-1997
		ES 2162211 T3	16-12-2001
		JP 10081693 A	31-03-1998
US 5846794 A	08-12-1998	FR 2749307 A1	05-12-1997
		AT 212383 T	15-02-2002
		CA 2206389 A1	29-11-1997
		DE 69709985 D1	14-03-2002
		EP 0810292 A1	03-12-1997
		JP 10087531 A	07-04-1998
WO 0114566 A	01-03-2001	AU 6571700 A	19-03-2001
		CA 2382462 A1	01-03-2001
		WO 0114566 A2	01-03-2001
		EP 1206551 A2	22-05-2002
		HU 0203049 A2	28-12-2002
		JP 2003507071 T	25-02-2003